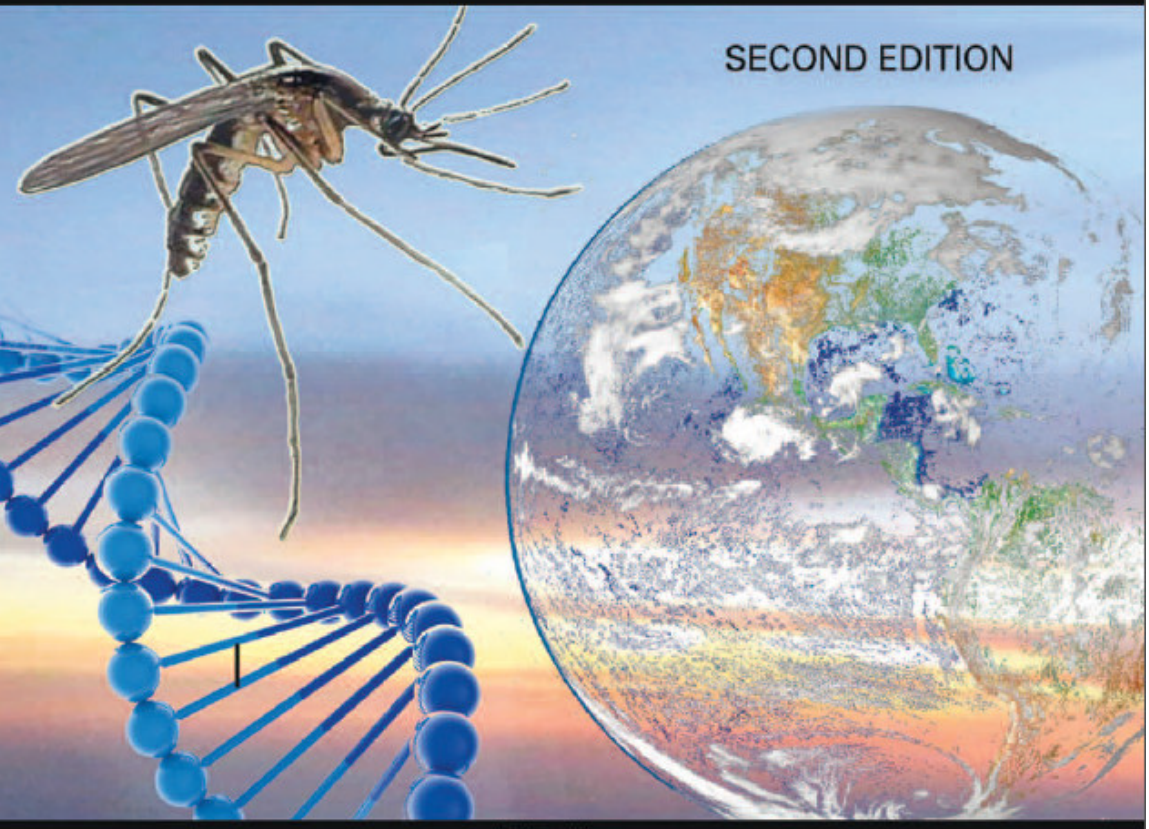
ELSEVIER

GENETICS AND EVOLUTION OF INFECTIOUS DISEASES

SECOND EDITION



Edited by

MICHEL TIBAYRENC



Genetics and Evolution of Infectious Diseases

Downloaded From Vet-eBooks.com

This page intentionally left blank

Downloaded From Vet-eBooks.com

Genetics and Evolution of Infectious Diseases

Second Edition

Michel Tibayrenc



ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Downloaded From Vet-eBooks.com

Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States

Copyright © 2017, 2011 Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-0-12-799942-5

For information on all Elsevier publications visit our website at https://www.elsevier.com/



Publisher: Sara Tenney
Acquisition Editor: Linda Versteeg-buschman
Editorial Project Manager: Halima Williams
Production Project Manager: Karen East and Kirsty Halterman
Cover Designer: Jenny Telleria

Typeset by TNQ Books and Journals

Downloaded From Vet-eBooks.com

Contents

List of Contributors **xiii**

1 Recent Developments in the Definition and Official Names of Virus Species **1**
M.H.V. Van Regenmortel
1. Introduction **1**
2. The Logic of Hierarchical Virus Classification **3**
3. Bionominalism: Are Species Classes or Individuals? **4**
4. The Virus Species Problem **5**
5. Properties Used for Defining Virus Species and Identifying Individual Viruses **8**
6. A Virus Species Cannot Be Defined Solely by the Properties of Viral Genomes **9**
7. The New ICTV Definition of Virus Species **11**
8. Non-Latinized Binomial Names for Virus Species **14**
9. Discussion **17**
References **18**

2 A Theory-Based Pragmatism for Discovering and Classifying Newly Divergent Species of Bacterial Pathogens **25**
F.M. Cohan, Sarah Kopac
1. Introduction **25**
2. Ecological Breadth of Recognized Species **30**
3. The Stable Ecotype Model of Bacterial Speciation **33**
4. Demarcating Putative Ecotypes From Sequence Data **35**
5. Ecological Diversity Within Putative Ecotypes **36**
6. Models of Frequent Speciation **38**
7. Other Models Where Ecotypes Are Not Discernible as Sequence Clusters **40**
8. Are Bacterial Ecotypes Cohesive? **41**
9. Incorporating Ecology Into Bacterial Systematics **43**
Acknowledgments **44**
References **44**

Downloaded From Vet-eBooks.com

3 Population Structure of Pathogenic Bacteria **51**
C.P. Andam, L. Challagundla, T. Azarian, W.P. Hanage, D.A. Robinson
1. Introduction **51**
2. Recombination in Bacterial Populations **52**
3. Evolutionary Processes Shape Intra- and Interhost Bacterial Population Structure **58**
4. Genomic Analysis Tools for Studying Bacterial Population Structure **61**
5. Conclusions **63**
References **64**

4 Epidemiology and Evolution of Fungal Pathogens in Plants and Animals **71**
P. Gladieux, E.J. Byrnes III, G. Aguileta, M. Fisher, R.B. Billmyre, J. Heitman, T. Giraud
1. Introduction **71**
2. New and Emerging Mycoses **74**
3. Plant Pathogenic Fungi **77**
4. New and Emerging Plant Diseases **78**
5. Modern Molecular Epidemiological Tools for Investigating Fungal Diseases **79**
6. Population Genetics of Pathogenic Fungi **80**
7. Genomics of Fungi: What Makes a Fungus Pathogenic? **86**
8. Conclusion **88**
References **88**

5 Clonal Evolution **99**
T. de Meeûs, F. Prugnolle
1. Introduction **99**
2. Definitions **99**
3. The Origin of Life, the Origin of Propagation, and Recombination **101**
4. Clonal Modes **102**
5. Quantifying the Importance of Asexuality in the Biosphere **103**
6. Genetic Consequences of Asexuality **103**
7. Evolution and the Paradox of Sex **105**
8. Clonal Microevolution **106**
9. Conclusions **109**
Abbreviation List **109**
Acknowledgments **110**
References **110**

6 Coevolution of Host and Pathogen **115**
A.D. Morgan, B. Koskella
1. Coevolution of Host and Pathogen **115**
2. The Process of Antagonistic Coevolution **121**

Downloaded From Vet-eBooks.com

3. Testing for Host–Pathogen Coevolution **124**
4. Implications of Coevolution **130**
5. Summary/Future Outlook **131**
References **132**

7 Microbes as Tracers of Past Human Demography and Migrations 141
J.-P. Rasigade, A. Gilabert, T. Wirth
1. Introduction **141**
2. Using Pathogens as Genetic Tracers for Host History **142**
3. Candidates **144**
4. Conclusion **159**
Abbreviations **159**
References **160**

8 Phylogenetic Analysis of Pathogens 167
D.A. Morrison
1. Introduction **167**
2. The Uses of Phylogenies **168**
3. The Logic of Phylogeny Reconstruction **169**
4. Characters and Samples **171**
5. The Practice of Phylogeny Reconstruction **174**
6. Choosing a Method **179**
7. Representing Phylogenies: Trees **181**
8. Phylogenetic Networks **184**
References **188**

9 Evolutionary Responses to Infectious Disease 195
G. Cochran, H. Harpending
1. Introduction **195**
2. Parasites as Our Friends **197**
3. Demography and Parasites **197**
4. Agriculture **198**
5. Some Lessons From Malaria **199**
6. Disease and Standard of Living in Preindustrial Societies: A Simple Model **202**
7. Population Limitation **203**
8. Disease, Mating, and Reproductive Strategy **206**
9. Prosperity and the Postindustrial Era Mortality Decline **207**
References **208**

10 Infectious Disease Genomics 211
Y.-T. Liu
1. Introduction **211**
2. Vaccine Target **214**
3. New Drug Discovery **214**

Downloaded From Vet-eBooks.com

4. Drug Target **215**
5. Therapeutic Response and Drug Resistance **216**
6. Vector Control **217**
7. Clinical Application **218**
8. Conclusion **219**
References **220**

11 Proteomics and Host–Pathogen Interactions: A Bright Future? 227
D.G. Biron, D. Nedelkov, D. Missé, P. Holzmuller
1. Introduction **227**
2. Interest of Proteomics to Study Host–Pathogen Interactions **228**
3. Retrospective Analysis of Previous Proteomics Studies **229**
4. Toward New Conceptual Approaches to Decipher the Host–Parasite Interactions for Parasites With Simple or Complex Life Cycle **236**
5. Population Proteomics: An Emerging Discipline to Study Host–Parasite Interactions **240**
6. Conclusion **247**
References **248**

12 The Evolution of Antibiotic Resistance 257
F. González-Candelas, I. Comas, J.L. Martínez, J.C. Galán, F. Baquero
1. Introduction **257**
2. Mechanisms and Sources of Antibiotic Resistance **260**
3. Evolution of Antibiotic-Resistance Genes **264**
4. Limitations to Adaptation and the Cost of Resistance **269**
5. Can the Evolution of Antibiotic Resistance be Predicted? **274**
6. Conclusions and Perspectives **275**
Glossary **276**
List of Abbreviations **276**
Acknowledgments **277**
References **277**

13 Modern Morphometrics of Medically Important Arthropods 285
J.-P. Dujardin
1. Introduction **285**
2. Landmark-Based Geometric Morphometry **285**
3. Pseudo-Landmark-Based Shape **288**
4. Allometry **289**
5. Measurement Error **289**
6. Some Considerations About the Genetics of Metric Change **290**
7. Phenotypic Plasticity **293**
8. A Special Case of Shape Change: the Character Displacement **293**
9. The Regulation of Phenotype **294**
10. Applications in Medical Entomology **296**

Downloaded From Vet-eBooks.com

Glossary 302
Acknowledgments 302
References 302

14 Evolution of Resistance to Insecticide in Disease Vectors 313
P. Labbé, J.-P. David, H. Alout, P. Milesi, L. Djogbénou, N. Pasteur, M. Weill
1. Introduction 313
2. Insecticide Resistance: Definition and History 314
3. Mechanisms of Resistance 318
4. Conclusion 328
References 329

15 Genetics of Major Insect Vectors 341
P.L. Dorn, S. Justi, E.S. Krafsur, G.C. Lanzaro, A.J. Cornel, Y. Lee, C.A. Hill
1. Introduction 341
2. Genetics of Tsetse Flies and African Trypanosomiasis 343
3. Genetics of the Triatominae (Hemiptera, Reduviidae) and Chagas Disease 351
4. The *Anopheles gambiae* Complex 361
5. Genetics of the Order Ixodida 367
Glossary 371
Acknowledgments 372
References 372

16 Multilocus Sequence Typing of Pathogens: Methods, Analyses, and Applications 383
M. Pérez-Losada, M. Arenas, E. Castro-Nallar
1. Introduction 383
2. Molecular Design and Development of Multilocus Sequence Typing 384
3. Multilocus Sequence Typing Databases 388
4. Advantages and Disadvantages of Multilocus Sequence Typing 389
5. Analytical Approaches 390
6. Applications of Multilocus Sequence Typing 395
7. Conclusions and Prospects 397
Acknowledgments 397
References 398

17 Next-Generation Sequencing, Bioinformatics, and Infectious Diseases 405
R. van Aerle, M. van der Giezen
1. Analyzing Big Data 405
2. Comparative Genomics 405
3. Transcriptomics 409

Downloaded From Vet-eBooks.com

4. Single-Cell Technologies **411**
5. High-Throughput Sequencing **413**
6. *De Novo* Genome Assembly **414**
7. Whole-Genome Sequence Analysis **414**
8. RNA-Seq (Transcriptomics) **415**
9. Concluding Remarks **416**
References **417**

18 Genomics of Infectious Diseases and Private Industry **421**
G. Vernet
1. Introduction **421**
2. Technologies and Instrument Platforms **423**
3. Customers and Their Needs **428**
4. Industry Landscape **430**
5. Conclusion **433**
References **433**

19 Current Progress in the Pharmacogenetics of Infectious Disease Therapy **435**
E. Elliot, T. Mahungu, A. Owen
1. Introduction **435**
2. Pharmacogenetics of HIV Therapy **435**
3. Pharmacogenetics of Antimalarial Therapy **442**
4. Pharmacogenetics of Antituberculous Therapy **444**
5. Summary and Perspective **446**
References **446**

20 Genetic Exchange in Trypanosomatids and Its Relevance to Epidemiology **459**
W. Gibson, M.D. Lewis, M. Yeo, M.A. Miles
1. Introduction **459**
2. *Trypanosoma brucei* **459**
3. *Trypanosoma cruzi* **465**
4. *Leishmania* **474**
Abbreviations **476**
Acknowledgments **477**
References **477**

21 Genomic Insights Into the Past, Current, and Future Evolution of Human Parasites of the Genus *Plasmodium* **487**
C.J. Sutherland, S.D. Polley
1. Introduction **487**
2. Evolution of *Plasmodium*: The Last 10 Million Years **491**
3. Evolution of *Plasmodium*: The 21st Century in Three Courses **496**
4. Evolution of *Plasmodium*, and the Eradication Agenda **501**
References **502**

Downloaded From Vet-eBooks.com

22 Integrated Genetic Epidemiology of Chagas Disease **509**
M. Tibayrenc, M.A. Shaw
1. What Is Integrated Genetic Epidemiology? **509**
2. Chagas Disease: A Major Health Problem in Latin America and Other Countries **509**
3. The Chagas Disease Cycle **510**
4. Host Genetic Susceptibility to Chagas Disease **510**
5. Vector Genetic Diversity **517**
6. Parasite Genetic Diversity **517**
7. Concluding Remarks **521**
Glossary **522**
References **522**

23 Adaptive Evolution of the *Mycobacterium tuberculosis* Complex to Different Hosts **529**
E. Broset, J. Gonzalo-Asensio
1. Overview: Disease and Mycobacterial Genetics **529**
2. Host−Pathogen Coevolution of the Tubercle Bacillus **531**
3. Evolution of the *Mycobacterium tuberculosis* Complex From a Genomic Perspective **538**
4. Evolution in the Laboratory Environment and In Vitro Attenuation of Bacteria From the *Mycobacterium tuberculosis* Complex **540**
5. Short-Term Evolution of *Mycobacterium tuberculosis* During Infection, Drug Treatment, and Disease **542**
6. Adaptive Cues of the *Mycobacterium tuberculosis* Complex As the Most Successful Pathogens **544**
7. Pending Questions and Concluding Remarks **547**
References **548**

24 The Evolution and Dynamics of Methicillin-Resistant *Staphylococcus aureus* **553**
M.M.H. Abdelbary, P. Basset, D.S. Blanc, E.J. Feil
1. Introduction **553**
2. The Staphylococcal Cassette Chromosome *mec* **553**
3. Evolution of *Staphylococcus aureus* and MRSA **556**
4. Molecular Epidemiology of MRSA **561**
5. Conclusion **564**
References **565**

25 Origin and Emergence of HIV/AIDS **573**
M. D'arc, L. Etienne, E. Delaporte, M. Peeters
1. History of AIDS **573**
2. Human Immunodeficiency Viruses Are Closely Related to Simian Immunodeficiency Virus From Nonhuman Primates **575**

Downloaded From Vet-eBooks.com

3. HIV-1 Is Derived From Simian Immunodeficiency Viruses Circulating Among African Apes **581**
4. Origin of HIV-2: Another Emergence, Another Epidemic **585**
5. Ongoing Exposure of Humans to a Large Diversity of Simian Immunodeficiency Viruses: Risk for a Novel HIV? **587**
6. Conclusion **591**
References **592**

26 Evolution of SARS Coronavirus and the Relevance of Modern Molecular Epidemiology **601**
Z. Shi, L.-F. Wang
1. A Brief History of SARS **601**
2. SARS Coronavirus **603**
3. The Animal Link **605**
4. Natural Reservoirs of SARS-CoV **606**
5. Molecular Evolution of SARS-CoV in Humans and Animals **610**
6. Coronavirus Surveillance in Wildlife Animals **614**
7. Concluding Remarks **614**
References **615**

27 Ecology and Evolution of Avian Influenza Viruses **621**
A.C. Hurt, R.A.M. Fouchier, D. Vijaykrishna
1. Introduction to Influenza A Virus **621**
2. Influenza Viruses in Birds **624**
3. Evolutionary Genetics of Avian Influenza Viruses **629**
4. Future Perspective **633**
Acknowledgments **634**
References **635**

Index **641**

Downloaded From Vet-eBooks.com

List of Contributors

M.M.H. Abdelbary Centre Hospitalier Universitaire Vaudois and University Hospital of Lausanne, Lausanne, Switzerland

G. Aguileta Ecologie Systematique Evolution, Univ. Paris-Sud, CNRS, AgroParis-Tech, Université Paris-Saclay, Orsay, France

H. Alout Institut des Sciences de l'Evolution de Montpellier (UMR 5554, CNRS-UM-IRD-EPHE), Université de Montpellier, Montpellier, France; Colorado State University, Fort Collins, CO, United States

C.P. Andam Harvard T. H. Chan School of Public Health, Boston, MA, United States

M. Arenas University of Porto, Porto, Portugal; Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP), Porto, Portugal

T. Azarian Harvard T. H. Chan School of Public Health, Boston, MA, United States

F. Baquero CIBER en Epidemiología y Salud Pública, Madrid, Spain; IRYCIS-Hospital Ramón y Cajal, Madrid, Spain

P. Basset Centre Hospitalier Universitaire Vaudois and University Hospital of Lausanne, Lausanne, Switzerland

R.B. Billmyre Duke University Medical Center, Durham, NC, United States

D.G. Biron Laboratoire Microorganismes: Génome et Environnement, UMR CNRS/UBP/UDA 6023, Aubière Cedex, France

D.S. Blanc Centre Hospitalier Universitaire Vaudois and University Hospital of Lausanne, Lausanne, Switzerland

E. Broset Universidad de Zaragoza, Zaragoza, Spain; CIBERes, Instituto de Salud Carlos III, Madrid, Spain

E.J. Byrnes III Duke University Medical Center, Durham, NC, United States

E. Castro-Nallar Universidad Andrés Bello, Santiago, Chile

L. Challagundla University of Mississippi Medical Center, Jackson, MS, United States

G. Cochran University of Utah, Salt Lake City, UT, United States

Downloaded From Vet-eBooks.com

F.M. Cohan Wesleyan University, Middletown, CT, United States

I. Comas FISABIO/CSISP-UV/Instituto Cavanilles, Valencia, Spain; CIBER en Epidemiología y Salud Pública, Madrid, Spain; IBV-CSIC, Valencia, Spain

A.J. Cornel University of California at Davis, Davis, CA, United States; Mosquito Control Research Lab, Parlier, CA, United States

M. D'arc UMI 233, Institut de Recherche pour le Développement (IRD), INSERM U1175 and Université de Montpellier, Montpellier, France; Instituto Nacional de Câncer, Rio de Janeiro, Brazil

J.-P. David Laboratoire d'Ecologie Alpine (UMR 5553 CNRS-UGA), Université Grenoble−Alpes, Grenoble, France

E. Delaporte UMI 233, Institut de Recherche pour le Développement (IRD), INSERM U1175 and Université de Montpellier, Montpellier, France

T. de Meeûs UMR 177 IRD − CIRAD INTERTRYP, Campus International de Baillarguet, Montpellier, France

L. Djogbénou Université d'Abomey Calavi, Cotonou, Benin

P.L. Dorn Loyola University New Orleans, New Orleans, LA, United States

J.-P. Dujardin CIRAD-IRD, Baillarguet, France

E. Elliot University of Liverpool, Liverpool, United Kingdom

L. Etienne UMI 233, Institut de Recherche pour le Développement (IRD), INSERM U1175 and Université de Montpellier, Montpellier, France; International Center for Infectiology Research, INSERM U1111, Ecole Normale Supérieure de Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5308, 69364, Lyon, France

E.J. Feil University of Bath, Bath, United Kingdom

M. Fisher Imperial College London, London, United Kingdom

R.A.M. Fouchier Erasmus Medical Center, Rotterdam, the Netherlands

J.C. Galán CIBER en Epidemiología y Salud Pública, Madrid, Spain; IRYCIS-Hospital Ramón y Cajal, Madrid, Spain

W. Gibson University of Bristol, Bristol, United Kingdom

A. Gilabert Laboratoire MIVEGEC, UMR 5290, IRD-CNRS-UM, Montpellier, France

T. Giraud Ecologie Systematique Evolution, Univ. Paris-Sud, CNRS, AgroParis-Tech, Université Paris-Saclay, Orsay, France

P. Gladieux Ecologie Systematique Evolution, Univ. Paris-Sud, CNRS, AgroParis-Tech, Université Paris-Saclay, Orsay, France

Downloaded From Vet-eBooks.com

F. González-Candelas FISABIO/CSISP-UV/Instituto Cavanilles, Valencia, Spain; CIBER en Epidemiología y Salud Pública, Madrid, Spain

J. Gonzalo-Asensio Universidad de Zaragoza, Zaragoza, Spain; CIBERes, Instituto de Salud Carlos III, Madrid, Spain; Hospital Universitario Miguel Servet, Zaragoza, Spain

W.P. Hanage Harvard T. H. Chan School of Public Health, Boston, MA, United States

H. Harpending University of Utah, Salt Lake City, UT, United States

J. Heitman Duke University Medical Center, Durham, NC, United States

C.A. Hill Purdue University, West Lafayette, IN, United States

P. Holzmuller UMR CIRAD-INRA Contrôle des maladies exotiques émergentes (CMAEE), Montpellier Cedex, France

A.C. Hurt WHO Collaborating Centre for Reference and Research on Influenza, VIDRL, at the Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia; University of Melbourne, Parkville, VIC, Australia

S. Justi University of Vermont, Burlington, VT, United States

Sarah Kopac University of Rhode Island, Kingston, RI, United States

B. Koskella University of California, Berkeley, CA, United States

E.S. Krafsur Iowa State University, Ames, IA, United States

P. Labbé Institut des Sciences de l'Evolution de Montpellier (UMR 5554, CNRS-UM-IRD-EPHE), Université de Montpellier, Montpellier, France

G.C. Lanzaro University of California at Davis, Davis, CA, United States

Y. Lee University of California at Davis, Davis, CA, United States

M.D. Lewis London School of Hygiene and Tropical Medicine (LSHTM), London, United Kingdom

Y.-T. Liu University of California San Diego, La Jolla, CA, United States

T. Mahungu Royal Free London NHS Foundation Trust, London, United Kingdom

J.L. Martínez CNB-CSIC, Madrid, Spain

M.A. Miles London School of Hygiene and Tropical Medicine (LSHTM), London, United Kingdom

P. Milesi Institut des Sciences de l'Evolution de Montpellier (UMR 5554, CNRS-UM-IRD-EPHE), Université de Montpellier, Montpellier, France

D. Missé Laboratoire MIVEGEC, UMR CNRS 5290/IRD 224/UM, Montpellier Cedex, France

Downloaded From Vet-eBooks.com

A.D. Morgan University of Edinburgh, Edinburgh, United Kingdom

D.A. Morrison Uppsala University, Uppsala, Sweden

D. Nedelkov Arizona State University, Tempe, AZ, United States

A. Owen University of Liverpool, Liverpool, United Kingdom

N. Pasteur Institut des Sciences de l'Evolution de Montpellier (UMR 5554, CNRS-UM-IRD-EPHE), Université de Montpellier, Montpellier, France

M. Peeters UMI 233, Institut de Recherche pour le Développement (IRD), INSERM U1175 and Université de Montpellier, Montpellier, France

M. Pérez-Losada George Washington University, Ashburn, VA, United States; Universidade do Porto, Vairão, Portugal; Children's National Medical Center, Washington, DC, United States

S.D. Polley Hospital for Tropical Diseases, London, United Kingdom

F. Prugnolle Maladies Infectieuses et Vecteurs Ecologie, Génétique, Evolution et Contrôle, MIVEGEC (IRD 224-CNRS 5290-UM1-UM2), IRD Center, Montpellier, France

J.-P. Rasigade Ecole Pratique des Hautes Etudes, Paris Sciences et Lettres, Paris, France; CIRI, International Center for Infectiology Research, INSERM U1111-CNRS UMR5308, ENS Lyon-Université Lyon 1, Hospices Civils de Lyon, Lyon, France

D.A. Robinson University of Mississippi Medical Center, Jackson, MS, United States

M.A. Shaw University of Leeds, St James's University Hospital, Leeds, United Kingdom

Z. Shi Chinese Academy of Sciences (CAS), Wuhan, China

C.J. Sutherland London School of Hygiene & Tropical Medicine, London, United Kingdom

M. Tibayrenc Maladies Infectieuses et Vecteurs Ecologie, Génétique, Evolution et Contrôle MIVEGEC (IRD 224-CNRS 5290-UM1-UM2), IRD, Montpellier, France

R. van Aerle Centre for Environment, Fisheries and Aquaculture Science, Weymouth, United Kingdom

M. van der Giezen University of Exeter, Exeter, United Kingdom

M.H.V. Van Regenmortel Université de Strasbourg-CNRS, Illkirch Cedex, France

G. Vernet Centre Pasteur, Yaoundé, Cameroon

D. Vijaykrishna Duke-NUS Graduate Medical School, Singapore, Singapore

L.-F. Wang Duke-NUS Medical School, Singapore, Singapore

Downloaded From Vet-eBooks.com

M. Weill Institut des Sciences de l'Evolution de Montpellier (UMR 5554, CNRS-UM-IRD-EPHE), Université de Montpellier, Montpellier, France

T. Wirth Ecole Pratique des Hautes Etudes, Paris Sciences et Lettres, Paris, France; Institut de Systématique, Evolution, Biodiversité, UMR-CNRS 7205, Muséum National d'Histoire Naturelle, Université Pierre et Marie Curie, Ecole Pratique des Hautes Etudes, Sorbonne Universités, Paris, France

M. Yeo London School of Hygiene and Tropical Medicine (LSHTM), London, United Kingdom

Downloaded From Vet-eBooks.com

This page intentionally left blank

Downloaded From Vet-eBooks.com

Recent Developments in the Definition and Official Names of Virus Species*

1

M.H.V. Van Regenmortel
Université de Strasbourg-CNRS, Illkirch Cedex, France

1. Introduction

Classification deals with abstract classes and taxonomy deals with classes called taxa. Viral taxonomy refers both to the scientific discipline of virus classification and to the outcome of a classification activity involving viruses.

Virus classification deals with abstract classes of viruses that are conceptual constructions of the mind. The most important characteristic of such classes is that they have members that are the concrete viral objects studied by virologists. Every membership condition determines a class and if a virus has a monopartite negative strand RNA genome, it automatically becomes a member of the *Mononegavirales*, which is a class known as an order.[1] Such a class is not physically real and must not be confused with the viruses themselves. Similarly, the abstract concept of a virus species as a class of viruses should not be confused with the viruses that are the concrete members of the species. Confusions between different logical categories have been a fertile source of misunderstandings in viral taxonomy. It has been claimed, for instance, that the name tobacco mosaic virus is an abstraction because only its particles can be handled.[2,3] Such a claim arose because the term "virus" was not recognized to be what logicians call a general term, that is, a word that denotes any number of concrete entities.[4(pp. 90–105)]

The species taxon was introduced in virus classification as late as 1991 when it was endorsed by the International Committee on Taxonomy of Viruses (ICTV), which is the body empowered by the International Union of Microbiological Societies (IUMS) to make decisions on matters of virus classification and nomenclature.[5] The official definition of virus species was as follows: "A virus species is a polythetic class of viruses that constitute a replicating lineage and occupy a particular ecological niche." Its key feature was that it incorporated the notion of polythetic class also known as a cluster class. While monothetic classes are defined by one or a few properties that are both necessary and sufficient for membership in the class, polythetic classes are defined by a variable set of statistically covariant properties, none of which is a defining property

* A very similar version of this chapter with the title: "Classes, taxa and categories in hierarchical virus classification: a review of current debates on definitions and names of virus species" appeared in Bionomina 2016, Dumerilia, 10: 1–20. copyright M. Van Regenmortel.

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00001-9
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

necessarily present in every member of the class. Since a virus species class is a conceptual construction, it cannot be *described* by its physical or material properties and can only be *defined* by listing certain properties of the viruses that are its members. Properties used for defining virus species are properties of viruses that can be altered by a few mutations, such as their natural host range, pathogenicity, mode of transmission, and small differences in the viral genome. This means that these species-defining properties vary considerably in different members of the same virus species. Since higher taxa, such as genera and families, have more viruses as members than species taxa, they require fewer defining properties than species taxa that need more properties to meet the qualifications for membership.

Section 2 clarifies the logical relations that exist among individual viruses, among the classes of the viruses called taxa and among the classes of the classes called categories. Since virus classification follows the structure of the Linnaean hierarchy, the logical structure of this hierarchy described by Buck and Hull[6] will be outlined.

The popular bionominalist school of thought, which claims that species are concrete individuals with a definite spatiotemporal localization instead of timeless, abstract classes, is examined in Section 3. The ontology advocated by Mahner and Bunge[7(pp. 253–70)] is described, leading to the conclusion that bionominalism does not provide an adequate framework for classifying viruses.

Section 4 reviews how the concept of virus species as a polythetic class finally became accepted in viral taxonomy, while Sections 5 and 6 clarify that because the properties used for demarcating individual virus species are easily modified by a few mutations, it is not possible to define virus species by relying only on one or a few necessary and sufficient conditions for establishing species membership. Although it is currently fashionable to employ small nucleotide motifs present in a viral genome as a DNA-barcoding system for identifying members of previously established viral taxa,[8] it must be emphasized that the presence of such motifs cannot be used by taxonomists as the sole defining criterion for creating or establishing new virus species.[9]

Section 7 describes the current debate surrounding the following controversial new definition of virus species ratified by the ICTV in 2013, which removed the term polythetic from the definition: "A virus species is a monophyletic group of viruses whose properties can be distinguished from those of other species by multiple criteria." Since every virus species, genus, or family could be considered to be a monophyletic group, this was actually a definition of "virus taxon" instead of virus species. A major concern was that "polythetic class" in the earlier definition had been replaced by "group" of viruses, because a group is a collection of viruses that are linked by a part–whole relation, whereas the term class implies the logical relations of class membership and class inclusion used in all hierarchical classifications.

Section 8 discusses the introduction of non-Latinized binomial names (NLBNs) for virus species. In 1998 the ICTV had introduced species names that differed from virus names only by typography, with the result that measles virus became officially a member of the species *Measles virus* (italicized, with a capital initial). This led to considerable confusion and the ICTV subsequently agreed that its Study Groups in charge of the taxonomy and nomenclature of individual virus families could propose NLBNs for species in certain virus genera. Such NLBNs, which had been used unofficially for

Downloaded From Vet-eBooks.com

50 years, are obtained by replacing the terminal word "virus" that occurs in all common English virus names with the name of the genus to which the virus belongs, which also ends in *-virus*. Measles virus thus became a member of the species *Measles morbillivirus*.

2. The Logic of Hierarchical Virus Classification

The root of the word classification is class, a term that refers to all the classes of viruses or organisms that have concrete objects as their members. Every membership condition determines a class, and since whatever is said about a thing ascribes a property to it, properties and classes are related entities.[10(pp. 22–4)] Bionominalists, however, deny that species are classes (see Section 3), although they consider genera and families to be classes since a taxonomy otherwise becomes impossible.[11]

Class membership is the logical relation that makes it possible to establish a bridge between two logical categories, namely an abstract class or taxon that is a mental construct and its concrete members that are objects located in space and time. This membership relation is different from the part–whole relationship that exists between two concrete objects, one being a part of the other in the way that a limb is part of a body. It is not possible for a viral object to be a part of a conceptual species construct, nor is it possible for a thought or concept to be part of a material object.[4(pp. 118–24)]

A virus is always a member of a certain virus species, which is the lowest taxon category in a hierarchical classification. Since a species taxon is also a class, it is a member of the species category which is the class of all species taxa. The relation that individual viruses have to taxa is the same membership relation that taxa have to their respective categories, the members of the category species being all the species taxa.[6]

Virus taxonomy makes use of a hierarchy of taxa, the lowest taxon being a virus species followed by the higher taxa of genera, families, and orders. The viruses that are members of a species taxon are also members of a genus taxon immediately above it as well as members of a higher family or order category. Similar species are collected in a genus, similar genera into a family, and similar families into an order.

The relation between a lower taxon and a higher taxon immediately above it is called "class inclusion," which is a crucial relation in the logic of a hierarchical classification. To say that the species *Measles virus* is included in the genus *Morbillivirus* is to say that the properties required for classifying a virus as a member of the species *Measles virus* include, besides others, all the properties required to classify it as a member of the genus *Morbillivirus*. The lower taxon, having fewer viruses as members, requires more properties to meet the qualifications for membership. This situation illustrates the logical principle according to which reducing the number of required qualifications increases membership whereas increasing the number of qualifications decreases membership.[6] The genus *Morbillivirus* can thus be regarded as a class generated by relaxing the membership requirement for being a member of the species *Measles virus*. Class inclusion in the Linnaean hierarchy obviates the need for

Downloaded From Vet-eBooks.com

repeating the properties used for defining higher taxa in the definitions of the lower taxa included in them, although all these properties are still necessary for membership in the lower taxon. It should be noted that this principle invalidates the claim that a single property may be sufficient for defining a virus species (see Sections 5 and 6). Higher taxa such as virus families and orders can usually be defined by a small number of stable, invariant properties that are both necessary and sufficient for membership in the class, which is the reason why these classes tend to be viewed as universal classes. Membership in such classes is thus easier to establish than membership in a virus species. It should be evident that the relation of class inclusion does not mean that the defining properties of a species are also automatically defining properties of a genus. The taxonomic categories of species, genera, families, and orders, which are classes of classes, are not included in each other since these categories are mutually exclusive classes. It is thus impossible, for instance, for the categories species and genus to have any taxa members in common.[6]

3. Bionominalism: Are Species Classes or Individuals?

The school of thought known as bionominalism considers that since species change during evolution, giving rise to new species, they must be evolving historical entities with a temporal dimension rather than immutable and timeless classes. This gave rise to the view that species are concrete individuals rather than abstract classes,[12,13] and changed the ontological status of species that no longer were considered to have viral objects as their members since viruses were now actually *part* of a material species. Mahner and Bunge[7(pp. 232–70)] analyzed in great detail the numerous consequences of this altered ontology of species-as-individuals (SAI). First, it is no longer possible to *define* species since only abstract concepts can be defined intensionally (see Section 5), with the result that only the proper names of taxa can be defined and not the taxa themselves.[6] Second, viruses are then linked to species taxa by part–whole relations instead of membership relations, which undermines the traditional view that classes, taxa, and any resulting classification are conceptual constructs rather than real objects. The SAI thesis also holds that species are lineages of ancestral-descendant populations with a spatiotemporal location and that all taxa are so-called historical entities forming cohesive wholes.[14,15] The notion of historical entity takes the history and lineage of a thing to be a concrete individual, which is an instance of reification. However, the history of a population is not a concrete system and the relation of antecedence is not a bonding but only a temporal relation. Mahner and Bunge[7(p. 238)] argued that the relational concepts of ancestry, progeny, and lineage are actually not real objects and that the so-called "genealogical nexus" is not a bonding or causal relation since the ancestry and progeny of a population cannot act upon each other unless they exist at the same time. Descent is not a causal relation since causality relates only events and not things, the caterpillar not being the cause of the butterfly. Species also cannot descend from each other in a literal sense since only concrete organisms or viruses

Downloaded From Vet-eBooks.com

can do so. When the only necessary and sufficient property for belonging to any taxon is descent from a common ancestor, it has been suggested[16] that descent may have become the new essence of the antiessentialists.

Since a classification is only a conceptual construct, taxa will be considered to be real individuals only when concepts are conflated with their referents. The major shortcoming of bionominalism is that it fails to distinguish between species as concrete entities and species as abstract entities, that is, it does not distinguish a thing from its conceptual representation. It is also possible to conceive and establish classes of objects that exist on Earth for only limited amounts of time, and species are clearly such classes. The class of paintings belonging to the French impressionist school is an example of a class with a historical dimension. As argued by Mahner and Bunge,[7] there is indeed good evidence that the mistaken ontology underlying bionominalism is responsible for its inability to provide an adequate philosophical framework for any biological classification.

4. The Virus Species Problem

The term species is used to denote the lowest category in a virus classification. Although viruses are not alive,[17] they are considered to belong to biology, and as such they are classified using the categories species, genus, family, and order employed in biology. In the case of genera and families, virologists readily accept that these categories are conceptual constructions of the mind, which should not be confused with real objects, since a virus family, for instance, cannot be purified, centrifuged, sequenced, or visualized in an electron microscope. Concepts such as virus species, on the other hand, are often viewed as more "real" than genera and families because they tend to be perceived as individual kinds of viruses infecting particular hosts. Some philosophers claim that concepts and objects can both "exist" because of the ambiguity of the term "exist."[4(p. 131)] The resulting confusion between species as an abstract class or category and species as concrete objects is common in the whole of biology (see Section 3), and attempts to resolve this confusion by devising a satisfactory definition of species is a problem that exists not only in virology. Darwin regarded the species category to be no more real than the categories genus and family, and his unwillingness to argue over the definition of species has been called a modern solution to the species problem.[18] However, the question whether species are real biological entities independent of any human conceptualization remains a hotly debated issue as illustrated, for instance, in the published exchange between Claridge[19] and Mishler.[20]

Biological species have been traditionally considered to be populations whose members can only breed among themselves and are reproductively isolated from those of other populations.[21] Since such a definition applies only to organisms that reproduce sexually, it was later modified to make it applicable to asexual organisms as follows: "A species is a reproductive community of populations, reproductively isolated from others, that occupies a specific niche in nature."[22]

Downloaded From Vet-eBooks.com

In the 1980s the view that there could be virus species was rejected by plant virologists because they assumed that the biological species concept of Mayr[21] defined by sexual reproduction, gene pools, and reproductive isolation was the only legitimate species concept, which obviously was not applicable to viruses that are replicated as clones.[23] Another reason why plant virologists were opposed to introducing species in virus classification was their belief that using the species category would bring about the use of Latin species names which they strongly opposed.[24] Although the first ICTV reports[25,26] advocated a Latinized viral nomenclature, Latinized virus species names were not introduced, and in the Fifth ICTV report[27] the rules regarding the use of Latin in virus taxonomy were removed, opening the way for the acceptance of virus species by plant virologists.

Another reason for the reluctance of many virologists to use the concept of species in virus classification was the absence of a virus species definition adopted by the ICTV. Many definitions had been proposed but none gained general acceptance. A popular textbook of plant virology[28] proposed that "virus species are strains whose properties are so similar that there seems little value in giving them separate names." This suggested that attributing names to virus species was the same activity as developing a taxonomy. Another definition proposed that "A virus species is a population of viruses sharing a pool of genes that is normally maintained distinct from gene pools of other viruses,"[29] which was also deemed unsatisfactory because many viruses are replicated entirely by clonal means and do not possess gene pools.

In 1989 the following definition was proposed: "A virus species is a polythetic class of viruses that constitute a replicating lineage and occupy a particular ecological niche."[30] This definition indicated that the members of a virus species are not simply phenetically similar objects devoid of a common origin but are collections of objects related by common descent. It also incorporated the notion of a shared ecological niche[31] used by Mayr in his species definition, which is a relational, functional property of an organism or a virus rather than a vacant space waiting to be occupied.[7(pp. 181–85)] However, the main novelty of the 1989 species definition was that it included the notion of polythetic class, which by then had become generally adopted by taxonomists.[32,13(pp. 178–80)] While monothetic classes are universal classes defined by one or a few properties that are both necessary and sufficient for membership in the class, polythetic classes are defined by a variable combination of properties, none of which is a defining property necessarily present in every member of the class. This means that (1) each member of a polythetic species shares a certain number of properties, (2) each property is present in a large but unspecified number of members, and (3) no property is necessarily present in all the members of the class and absent in the members of other classes (Fig. 1.1). It should be stressed that the term polythetic only describes a particular distribution of properties present in a class and that the members of a class do not themselves possess polythetic or monothetic properties (see Section 7). Likewise, being a genetic parasite or having a vector is a property of viruses and not of classes. A concept such as a species class cannot have physical or material properties since only its members do. This means that species cannot be *described* but can only be *defined* by listing certain properties of their members. The viral objects that

Downloaded From Vet-eBooks.com

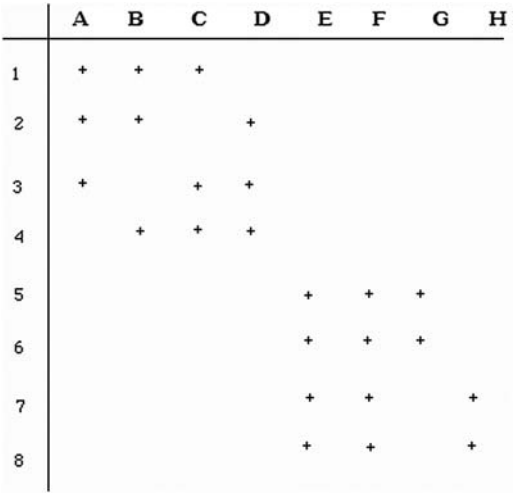


Figure 1.1 Distinction between polythetic and monothetic classes in the case of eight individuals (1−8) and eight properties (A−H). The possession of a property is indicated by a plus sign. Individuals 1−4 constitute a polythetic class, each member possessing three of four properties with no common property being present in all the members. Individuals 5−6 and 7−8 form two monothetic classes with three properties present in all the members
From Van Rijsbergen K. *Information retrieval*. 2nd ed. London: Butterworths; 1979; see also http://www.iva.dk/bh/lifeboat_ko/CONCEPTS/monothetic.htm.

the concept refers to are thus the concrete instances that satisfy the membership conditions of the class.

In 1991 the definition of species as a polythetic class was endorsed by the ICTV, and species became the lowest level in virus classification.[5] Unfortunately, many virologists thought that this definition would provide them with guidelines for establishing and demarcating new virus species and for deciding whether a virus was a member of a particular species. This misunderstanding led to a never ending debate about the presumed usefulness of a species definition for creating new species taxa and identifying their members (see Sections 5 and 7).

The concept of polythetic class, also known as a cluster class, is well established in taxonomy[13,32,33] and was used successfully by ICTV Study Groups for establishing many new virus species.[34] It has also been repeatedly emphasized that the members of a polythetic virus species possess a consensus set of statistically covariant properties but not a single, common property that is necessary and sufficient for membership in the class.[3,35−39] Nevertheless, a few virologists objected to species being called polythetic classes because they claimed that the term polythetic, commonly used in taxonomy, was obscure and not widely understood. It was also claimed that virus species could be defined monothetically by features of their gene sequences, and this led to the proposal that the term polythetic should be removed from the species definition.[40] Although there was considerable opposition to such a change,[41] the ICTV in 2013 introduced a new definition of virus species that no longer included the notion of polythetic class (see Section 7).

Downloaded From Vet-eBooks.com

5. Properties Used for Defining Virus Species and Identifying Individual Viruses

Properties are possessed by things and objects and cannot be detached from them. Intrinsic properties (or characters), such as chemical composition, are possessed regardless of other things whereas relational properties, such as being a genetic parasite or being vector-borne, are possessed by virtue of the relation of a virus to other things, such as a host or a vector.

Some philosophers distinguish a property from a predicate or attribute that they view as the conceptual representation of a thing's property. This distinction is important because not all predicates represent properties of real things. A thing either possesses property P or does not possess it, but it cannot possess the property "not P" since there are no negative properties. However, for every predicate, there is another predicate that is the negation of the first. Negation for instance may affect the proposition "tapeworms think" but not the property of thinking.[7(p. 10)]

The terms character, feature, and trait are often used in the sense of both property and part although a part of a thing is a thing and not a property.[42] The notion of character has been called the central mystery of taxonomy,[43] and although the possession of a certain part can be viewed as a property, it is not clear if a complete genome sequence, a particular nucleotide motif, or the presence of a certain nucleotide in a viral genome should count as a single character.[8,37,44]

Classes and concepts can only be defined whereas their individual members or any other concrete objects can only be described. Taxa are defined intensionally by sets of properties that provide the qualifications for membership in the class. The intension of a concept such as class is its meaning that, however, does not give it any reality outside the realm of intellectual constructions. The extension of the class is the set of members of the class, for instance the real viruses it refers to, which are the concrete referents of the class. Since the intension of a class determines its extension, the extension of a class can only be determined if one can distinguish members from nonmembers, which means that intension must precede extension.[7(p. 227)] A species taxon must first be established and defined by taxonomists before it becomes possible to ascertain if a sufficient number of the species-defining properties are present in an individual virus to make it a member of the species. The proposal that a monothetic (instead of a polythetic) species class can be established by relying on a single defining property such as a particular nucleotide motif found in viral genomes[40] overlooks the fact that it would be necessary to know beforehand that this motif is present in all the members of the species and absent in other species, which means that the extension would need to precede the intension.[9,44]

Properties useful for distinguishing individual species within a genus obviously cannot be the stable and invariant properties used for defining genera (such as the method of virus replication or the morphology of virus particles) that are the same in all the members of the genus. Properties used for defining virus species are properties of viruses that can be altered by a few mutations, such as natural host range, cell

Downloaded From Vet-eBooks.com

and tissue tropism, pathogenicity, mode of transmission, small genome differences, and so on. Many species-defining properties, therefore, tend to vary considerably in different members of the same species, which is the very reason why species are defined polythetically by a variable combination of properties. Since they are defined by an indefinite number of statistically covariant properties, species are considered the best examples of cluster classes.[13,45] A cluster class is defined by a cluster of properties, the majority of which may be present in all members of the class although some properties can be absent in individual members. Since all the species-defining properties are not necessarily present in every member of a species, viral taxonomists may have to create species by drawing boundaries across a continuum of phenotypic and genomic variability which often involves a strong subjective element.[37,38]

The continuous nature of biological variation often leads to an absence of clear-cut discontinuities between closely related species, which could then be considered as fuzzy classes with blurred boundaries. However, this would not justify abandoning the species concept since the continuous nature of electromagnetic radiation or of geological formations does not prevent us from recognizing different colors or individual mountains.[4(p. 125),37]

The demarcation of a species taxon by a virologist using the polythetic criterion should not be confused with the task of identifying a virus isolate as a member of a species. Once a species taxon has been established, it becomes possible to compare the properties of putative members of the species in order to discover one or more so-called diagnostic properties[12,36,46] that may suffice to identify the virus. Such diagnostic markers could be a specific reaction with a monoclonal antibody[47] or a particular nucleotide motif,[8] although these are not properties that are used to define a species taxon beforehand. The technique known as DNA barcoding[48] is sometimes presented as providing a useful additional character for producing new species although it is only a tool for identifying members of existing species. Nucleotide motifs cannot be used for distinguishing and establishing new species amid the thousands or millions of species that have not yet been sequenced or recognized on the basis of phenotypic or other criteria.[9,49]

6. A Virus Species Cannot Be Defined Solely by the Properties of Viral Genomes

It is now commonly accepted that virus classification should reflect the phylogenetic relationships among viruses, which can be established from the sequence divergence observed in viral genomes.[50,51] As more sequences of viral genomes became available, attempts were made to establish species only on the basis of genome data obtained from putative members of a viral species. As explained in Section 3, this cannot succeed since it is not possible to derive the intensional definition of a species from its extension. The DNA or RNA sequence present in a virion is part of the phenotype of the virus since it is a part of the virion chemical structure. Phenotypic properties

Downloaded From Vet-eBooks.com

include the morphology and molecular composition of the virion as well as the biochemical activities of the virus and all its relational interactions with hosts and vectors. A virus classification based on nucleotide sequences present in a virion is thus a phenotypic classification based solely on molecular sequences rather than on biological and functional properties.[7(p. 287)] There is no reason to assume that when virus species are demarcated only on the basis of genome sequences and a derived hypothetical phylogeny, this will necessarily produce a classification that is more correct, relevant, or useful than a classification based on all the phenotypic properties of a virus.[52] Genome characteristics do not by themselves justify taxonomic allocations, and the wish to record phylogeny should not overshadow the importance of other phenotypic and biological properties, which are the main reasons why virologists classify viruses and engage in species demarcation.

It is impossible to infer the total phenotype of a virus from its genotype because a phenotype is not simply the manifestation or expression of a genotype, but depends also on numerous contributions of extraneous epigenetic factors present in the environment and in viral hosts and vectors. This makes the phenotype the result of an ontogenic development involving both genetic and nongenetic factors.[53–55] It is sometimes claimed that most, if not all, biological properties of a virus could, at least in theory, be deduced from the sequences of its viral genome and encoded proteins. This is actually not the case since it is impossible, for instance, to predict from the sequence of encoded viral proteins which receptors of a virus determine its host and tissue specificity, as this would require prior knowledge of which host and tissues the virus is able to infect. The receptor-binding site of a virus is a relational structure existing by virtue of a relation with cellular receptors in the infected host. It is equally impossible to deduce the immunological properties of a virus or to predict how the immune system of a host is likely to react to a viral infection simply by predicting the presence of certain conformational epitopes in an encoded viral protein using ineffectual bioinformatic algorithms.[56]

In his analysis of the relationship between a unit of genotype that is genetically expressed and a unit of phenotype, Moss[57,58] argued that the metaphor of a gene as a code and information carrier arose from a conflation of two distinct meanings of the term gene that he called Gene-P and Gene-D. The Gene-P is defined by its relationship to a particular phenotypic character but does not entail the presence of a specific nucleic acid sequence able to initiate a series of developmental steps leading eventually to the phenotype. The classic example of this is the elusive Gene-P for blue eyes where the blue color results from the absence of a DNA sequence necessary for making a brown eye pigment. There may be many structural reasons for the absence of such a sequence and any one of them could count as a genetic factor for blue eyes. Speaking of a gene in the sense of Gene-P is sometimes useful because it allows predictive talk about the likelihood of some phenotypic property.[58(p. 44)]

A Gene-D, on the other hand, is defined by its molecular sequence and is a developmental resource (hence the “D”) which, however, cannot on its own determine the phenotype. A Gene-D does not specify the numerous transcriptional complexes that may result from differential RNA splicing nor all the intermediate products needed to achieve the ultimate phenotypic outcome. Phenotypes are achieved through the

Downloaded From Vet-eBooks.com

complex interaction of many factors, and Gene-D sequences are not adequate substitutes for other phenotypic properties.

When the concepts of Gene-P and Gene-D are conflated, it may give the impression that the entire chain of reactions that lead from transcriptional units to a phenotype has been elucidated although this is not the case. In fact, Gene-P is only a theoretical predictor device for some phenotype while Gene-D sequences do not specify all the developmental steps involved in producing the phenotype.

Taxa produced only on the basis of genome sequences do not necessarily agree with taxa established using additional biological and structural properties of viruses. It has been shown, for instance, that a classification based on both genome sequences and structural phenotypes may reveal additional evolutionary connections that are not detectable when only sequence-based approaches are used.[59,60]

The fact that the presence of a characteristic, short nucleotide motif in a virus isolate may suffice for identifying the isolate as a member of a particular virus species is sometimes regarded as evidence that the biological properties of a virus can be deduced from its genome sequence. However, being a member of a species does not imply that all the biological properties of the virus are firmly established since a virus species is a polythetic class of viruses defined by a variable combination of properties, none of which is necessarily present in all the members of the species. The diagnostic nucleotide motif possesses no causal efficacy in determining any particular biological property and if mutations are introduced in the motif, they are unlikely to reveal that the motif is responsible for a particular phenotype, since the function of a gene is not to produce whatever the system fails to do when the gene is absent or has been modified.[61] Predicting the host or vector specificity of a virus from its genome sequence always remains a hazardous enterprise because of our profound ignorance of how viral attachment proteins determine vector specificity or viral host range by controlling cell and tissue tropism during viral infection.[62]

In conclusion, it should be evident that if only sequences of viral genomes are taken into account, this will produce a classification of viral genomes rather than a classification of viruses. Viruses should not be reduced to sequences, and there is no justification for the claim that a genome-based classification that privileges phylogenetic considerations makes it superfluous to utilize all the known discriminating phenotypic properties of viruses for establishing species and other virus taxa that are useful to laboratory virologists.

7. The New ICTV Definition of Virus Species

In 2004 it was reported that all the RNA genome sequences of viruses belonging to the species *Tobacco mosaic virus* possessed a unique nucleotide combination motif (NC-motif) of 47 nucleotides, present in the viral polymerase gene. This NC-motif could be used for identifying all the members of that species and for distinguishing them from members of other species in the *Tobamovirus* genus.[8] Other NC-motifs

Downloaded From Vet-eBooks.com

were also found to be diagnostic markers for identifying viruses assigned to other species in the *Tobamovirus* genus, and one NC-motif was found that could identify any member of that genus. These findings led Gibbs and Gibbs[40] to propose that a virus species could be defined monothetically by the presence in all the members of the species of a common NC-motif, which they considered to be a species-defining property that was both necessary and sufficient to establish membership in the species. They removed therefore the term polythetic from the ICTV definition in use since 1991 and proposed the following so-called "broader" definition: "A virus species is a class of viruses that constitutes a replicating lineage and occupies a particular ecological niche." This was presented as the intentional meaning or definition of the concept of species class, based on the assumption that a part of a viral genome is a monothetic property, necessarily present in every member of the class. This implied that the extension of the class had to be known beforehand (see Section 5). The authors removed the term polythetic from the initial definition because they thought it meant a type of variable property rather than a certain distribution of properties (Section 4). When the proposal was posted on the ICTV website, it elicited unfavorable comments and it was subsequently not approved by the ICTV.

In July 2012, another proposal by four members of the ICTV Executive Committee, A. King, M. Adams, E. Lefkowitz, and E. Carstens, was posted on the ICTV website,[63] which included the following new definition of virus species: "A species is a monophyletic group of viruses whose properties can be distinguished from those of other species by multiple criteria." The authors acknowledged that these criteria could be genome properties and any other phenotypic properties of viruses, but they no longer included the requirement that the viruses had to form a polythetic group characterized by the absence of a single defining property necessarily present in all the members of the species. As a result, it became possible to establish species as monothetic groups of viruses that shared only one or a few common defining properties. For instance, if two anelloviruses possessed only 65% nucleotide identity in their genomes, this sole criterion was used to allocate them to two different species even in the absence of known differences in other biological or phenotypic properties. Furthermore, since every species, genus, and family can be considered to be a monophyletic group, King et al.[63] had in fact coined a definition of virus taxon instead of virus species. Another reason why monophyly is not a valid criterion for species demarcation is the common occurrence in many viruses of recombination and reassortment among parts of viral genomes, which produces chimeric viruses with polyphyletic genomes.[52] This makes it impossible to accurately represent such multidimensional phylogeny in a monophyletic scheme.[34] Many comments were posted on the ICTV website[64] opposing the proposed new definition of virus species. Subsequently, the proposers of the new species definition responded on the ICTV website with a 4-page, polemical document in which they claimed that the ICTV definition of virus species used since 1991 was based on specious reasoning and on meaningless terms, such as polythetic class, replicating lineage, and ecological niche. All the arguments and counterarguments are available on the ICTV website,[64] and the shortcomings of the proposed new

Downloaded From Vet-eBooks.com

definition were described in detail by a group of six ICTV Life Members and eight other senior virologists.[41] These shortcomings are summarized as follows:

1. In the ICTV forum,[64] the proposers of the new species definition repeated the same mistake as Gibbs and Gibbs[40] when they claimed that since a species corresponds to a replicating lineage, it cannot be a polythetic class because all its members must have inherited one or more properties from a common ancestor which makes the class a monothetic one. However, a variable distribution of properties in the members of the class together with the absence of a single common defining property in all of them is what defines a polythetic class. This combination of properties does not itself constitute a single common property shared by all the members of the class since it is a characteristic of the class rather than a property of its members. Every membership condition determines a class but since a class is not a concrete object, it cannot itself figure as a candidate for membership of the class. Virus classes only admit viruses as members but cannot admit themselves as members. Membership of classes can thus be determined by one or many membership conditions (e.g., virus properties), except one which is nonself membership.[65(p. 44),10(p. 94)] The adjective "long," for instance, denotes the class of long things but since it is not a long adjective, it is a nonself-denoting property of the class. If one fails to appreciate that the nonself-membership condition does not determine the class, one lands with the well-known Russell paradox of the barber,[41] which can be stated as follows. If one assumes that a village barber shaves all and only those men in the village who do not shave themselves, one lands with a contradiction since the barber will need to shave himself only if he does not do so.[66] The paradox is resolved only when it is realized that there is no such barber.[10(p. 146)] As clearly stated by Quine[10(p. 227)]: "When we say of some class that it is not a member of itself we do not thereby assign it to a class of all nonself members; for that class, if it existed, would have to be a member of itself if and only if it was not. Similarly when we say of some property that it is not a property of itself we do not thereby ascribe a property to it." The nonself-membership condition also excludes the possibility that the class of all polythetic classes could form a monothetic class that would have viruses (instead of classes) as its members.
2. The preposterous claim was also made by the proposers of the new definition[64] that the term "class" should only be used to denote a category in the classification hierarchy, that is, the one situated above the category order and below the category phylum, although such a category is not used in virus classification. They seem to be unaware that the conceptual construct of class is universally used in taxonomy because it makes it possible to establish a link between the abstract class and the concrete organisms or viruses that are members of a class. They object to a virus species being called a polythetic class and propose instead to define a species as a "group" of viruses. However, a group of viruses is only a collection of viral objects that are linked to the group by the part−whole relation, and such a terminology contradicts the logic of classes used in taxonomy that uses only the relations of class membership and class inclusion for building up a classification (see Section 3).
3. The proposers of the new definition[64] also dismissed the glaring case of the 178 begomovirus species that were created by ignoring the polythetic principle and accepting that species could be established on the basis of a single arbitrary criterion, namely less than 89% pairwise sequence identity in the viral DNA-A genome.[67] Virus classification by Pairwise Sequence Comparison (PASC) of viral genome sequences has increasingly been used.[68] It produces plots of the frequency distribution of pairwise identity percentages from all available genome sequences of viruses in a family that show multimodal distributions of peaks that may be tentatively attributed to clusters of sequences corresponding to groups of serotypes, strains, species, and genera.[37] Percentage demarcation thresholds used to allocate various peaks to

Downloaded From Vet-eBooks.com

different taxon categories could then be used, for instance, to decide how many species were present in a genus. In the case of the begomoviruses in the *Geminiviridae* family, a cut-off point of less than 89% sequence identity in the DNA-A genome was chosen as the sole criterion for separating strains from species, and this led to the creation of 178 different species in the absence of any biological evidence that such taxa corresponded to distinguishable, stable entities that would justify the label "species." It was argued[64] that the sole use of the 89% cut-off point was not an arbitrary decision because it was based on quantitative genome data. They refused to admit that choosing a lower percentage cut-off point in the PASC peaks could have produced a smaller, more reasonable number of begomovirus species. Many of the so-called 178 different "species" consist of viruses that infect the same host (cotton or tomato) and produce very similar disease symptoms, and had to be given different names by including the geographical location of the first isolation of the virus. This produced a long list of species names, such as *Tomato leaf curlComoros virus, Tomato leaf curl Guangxi virus, Tomato leaf curl, Hsinchu virus, Tomato leaf curl New Delhi*, and so on, which could have been considered strains of the same species if a lower threshold demarcation percentage for creating species had been chosen.[9] The allocation of different begomoviruses to the category strain or variant is equally arbitrary[69] while attributing a peak to the so-called "virus isolates" is in fact meaningless since isolates can refer to any virus that is being studied experimentally, which could be a member of a strain, species, or genus.[37] In the genus *Mastrevirus* in the same *Geminiviridae* family, a more appropriate cut-off point of 75% sequence identity in DNA-A sequences was used that led to the creation of only 12 separate species.[70] Establishing valid demarcation criteria in the family *Geminiviridae* is particularly difficult because of the frequent occurrence of recombination events between different geminiviruses.[34,71]

All previously mentioned objections to the new species definition were ignored by the ICTV EC, and the proposal was ratified using a fast-track approval process, which considerably reduced the time available for posting further objections and comments on the ICTV website. The ballot return rate of votes was 41% of those entitled to vote, with 45 in favor, 21 against, and 2 abstentions.[72] According to Van Regenmortel et al.[41] the new species definition is in no way superior, and in many ways inferior, to the earlier ICTV definition; by removing the polythetic principle, it certainly will not make it easier for virologists to establish or recognize new species in the future.

8. Non-Latinized Binomial Names for Virus Species

The assignment of names to virus taxa is the responsibility of the ICTV of IUMS. The names of virus genera, subfamilies, families, and orders have for many years been written in italics with a capital letter which is a different typography from that advocated for such taxa by the *Biological Code of Nomenclature*.[73] The unique position of viruses in biology is one of the reasons why the traditions of the *International Code of Zoological Nomenclature*,[74] the *International Code of Nomenclature of Bacteria*,[75] and the *International Code of Botanical Nomenclature*[76] are not followed by virologists. The ICTV as the voice of the international community of virologists has always followed its own rules and *Code* and tends not to follow traditions present in the rest of

Downloaded From Vet-eBooks.com

biology, such as the use of Latin names[24] or the formation of binomial species names using the order genus-name-first/species identifier-second, instead of the reversed order (species-first/genus-second) introduced in virology 50 years ago.[26] ICTV activities are governed by Statutes and by the *International Code of Virus Classification and Nomenclature*. The most recent version of the *Code* was published in the 9th ICTV Report in 2012.[77] Between successive Reports, ratified changes to the *Code* and to virus classification and nomenclature are brought to the attention of virologists in the "Virology Division News" (VDN) columns of *Archives of Virology*, the official journal of the Virology Division of IUMS.[78,79]

The respective names of virus taxa have the following endings: *virus* for genera, *-virinae* for subfamilies, *-viridae* for families, and *-virales* for orders.[1] It has been suggested that unofficial vernacular names for the members of these four taxa could be introduced using taxon-specific suffixes.[80] The suffixes are "-virad" for members of an order, "-virid" for members of a family, "-virin" for members of a subfamily, and "-genus" for members of a genus. This is useful, for instance, when a genus name such as *Parvovirus* served as a basis for coining the family (*Parvoviridae*) and subfamily (*Parvovirinae*) names. When referring to a parvovirus, it is not clear if one is thinking of a member of the family, subfamily, or genus, whereas referring to a parvovirid, parvovirin, or parvovirus removes any ambiguity.

Regarding species names, I had proposed to the ICTV Executive Committee (EC) in 1998 that two alternative changes could be introduced in species names. One proposal was to adopt the common English names of viruses as species names, but to italicize them with the initial letter capitalized in order to provide a visible sign that species correspond to taxonomic classes, just like italicized genera and families. The other proposal was to adopt NLBNs, which had been used unofficially for many years in plant virology papers and books[81–84] and in the indices of earlier ICTV Reports.[26,85,86] In the 5th ICTV Report,[27] NLBNs were retained only for indexing plant viruses and in the 6th Report,[87] they were dropped altogether because some animal virologists were opposed to their use. One argument against the introduction of binomial species names was that long-established virus names would have to be abandoned. However, this is not the case since original names of viruses would be retained, and the new names concerned only virus species for which names did not yet exist.

It was proposed in 1998 that NLBNs for species would be italicized with a capital initial and would be obtained by replacing the terminal word "virus" occurring in all common English virus names with the genus name to which the virus belongs, which also ends in *-virus*. Such a system would not require the creation of completely new names for thousands of virus species, which would be the case if Latin binomial names were introduced.[2,36,37,88,89]

In 1998 the majority of the members of the ICTV EC, who were not plant virologists, adopted the first proposal in spite of the fact that NLBNs could have been immediately endorsed for more than 90% of the 1550 virus species recognized at the time.[90] As a result of this decision, measles virus became officially a member of the species *Measles virus*.[91] Within a few years after the adoption of species names that differed from virus names only by typography, it became clear that many virologists found it difficult to use these names correctly because they constantly had to decide whether

Downloaded From Vet-eBooks.com

they wanted to refer to the virus or to the taxonomic species class, a distinction that many of them found difficult to make.[37,92–94] Virologists would, for instance, frequently write that *Measles virus* or *Cucumber mosaic virus* had been isolated, transmitted to a host, or sequenced although species, being taxonomic constructs of the mind, cannot have hosts, vectors, or sequences. Such logically incorrect statements are common in biology because the majority of animals, plants, and microorganisms have no vernacular names in English or other languages. Scientists will therefore write that *Escherichia coli* (i.e., the species) has been infected by a virus, as if a taxonomic concept could be infected.[95] In virology, such statements are easily avoided since all viruses have vernacular names and the name of the virus, instead of the species, can always be used to refer to the infectious agent. It has been suggested that introducing binomial species names should be postponed until laboratory virologists had fully grasped the nonidentity of conceptual species and concrete viruses.[94] This may well be counterproductive since it is actually by using species NLBNs that clearly differ from virus names that virologists would demonstrate in their writing that they understood the distinction. How else would one know that they had grasped it?

Virologists have come to realize that the use of species NLBNs has the advantage that because binomial names in biology are always associated with taxonomic entities, this makes it easier for them to recognize that binomial names are the names of virus species rather than of viruses.[96] It is also evident that NLBNs provide useful additional information on the properties of the viruses, deduced from membership in a genus, which was the reason Fenner started to use binomial names already in 1976.[26]

In 2002 efforts were made to canvass the opinion of virologists who attended an international Virology conference in Paris regarding their acceptance of species NLBNs. The results of two ballots showed that a significant majority (80–85%) of the 250 virologists who expressed an opinion were in favor of binomial names for species.[90,97] In 2004 half the members of the ICTV EC no longer objected to such names although the EC found it difficult to canvass the opinion of the more than 80 ICTV Study Groups because only a few of them made their views known.[98]

As it became obvious that NLBNs were superior to the official species names, a proposal was made to generalize the use of such binomial names for all virus species.[99] However, the ICTV EC decided that the use of such species names should not be mandatory, but that it should be left to Study Groups to initiate formal proposals if they wished to introduce binomial names for certain virus families. Jens Kuhn who is a member of several Study Groups as well as the editor responsible for the VDN section of *Archives of Virology* has been very active in introducing binomial species names in several families, such as the *Arenaviridae*, *Bornaviridae*, *Filoviridae*, *Nyamiviridae*, *Rhabdoviridae*, *Bunyaviridae*, and *Paramyxoviridae*.[79,100] Paradoxically, some plant virologists who had strongly criticized the ICTV in the past for not ratifying NLBNs for all virus species[101] became adepts of Latinized, binomial species names because they believed that viruses are organisms.[102] This belief goes against the well-established consensus that living organisms possess an autonomy and many metabolic and functional capacities that are never found in viruses or in any nonliving matter.[7(pp. 141–6),17,33]

Downloaded From Vet-eBooks.com

9. Discussion

Sections 4–8 of this review followed a chronological presentation of developments in viral taxonomy, which demonstrate that the field has been plagued by a continuous series of conflicting views, heated disagreements, and acrimonious controversies that may seem to some to be out of place in a scientific debate. The reason, of course, is that the subject of virus taxonomy and nomenclature lies at the interface between virological science and areas of philosophy, such as logic, ontology, and epistemology, which unfortunately are rarely taught in university curricula followed by science students.[103] Richard Feynman quipped that "philosophy of science is about as useful to scientists as ornithology is to birds" while Imre Lakatos lamented that: "most scientists tend to understand little more about science than fish about hydrodynamics."[104(p. 2)] It is indeed regrettable that a highly informative book such as the *Foundations of Biophilosophy*[7] does not feature more often as compulsory reading in postgraduate courses offered to biology students.

Philosophy abounds with contradictory views and interpretations regarding the nature of biological phenomena and the ongoing debate about species being classes or individuals (Section 3) is clearly a philosophical issue. What is more unexpected is that plant virologists were much more reluctant than animal virologists to accept virus species as useful classes in viral taxonomy and that they claimed that establishing such taxa "logically" entailed that they would be given Latin names, which they strongly opposed.[24]

The appeal to logic in such debates[2] is indeed astonishing since Latinization is only a matter of linguistic convention and tradition in biology, and most virologists do not view viruses as living organisms[17] that should be classified according to the rules of the proposed *Biocode*.[105] When virus species names eventually became italicized English binomial names instead of italicized Latin binomial names, Gibbs,[101,102] who claims that "plant virologists have a greater call on nomenclature than most working animal virologists," tried to downplay the contemporary primacy of English in virological communications by stating that Latin, anyway, had never been the language of communication between scientists, a claim that is patently untrue.[3] However, it cannot be denied that English has now replaced Latin as the predominant communication language used by scientists. The major journals and reference books in Virology are written in English and virologists, irrespective of their mother tongue, are familiar with English virus names. Inventing thousands of new Latin binomial names for virus species is unlikely to be a welcome alternative.

Claims that the ICTV is leading virus nomenclature into chaos have often been refuted,[106] and the derogatory tone that is sometimes used in such attacks has been deplored. There is indeed no ground for claiming that ICTV is breaking its own rules since it only amends them following due process, or for asserting that ICTV has become isolated from its broader electorate of virologists and no longer represents their interests.[102] ICTV activities are increasingly displayed in the VDN columns[78,79,107] and the advice extended by Gibbs that all virologists should ignore the ICTV is itself a neat recipe for chaos and is best dismissed as provocation.[106] The ICTV has also

Downloaded From Vet-eBooks.com

been criticized for not providing extensive descriptions of individual viruses in their ICT Reports. This task was supposed to be fulfilled by the Universal Virus Database,[108] and it is unfortunate that this project has now been abandoned.[34,109]

ICTV is a democratic organization and it has refused to implement a mandatory system of NLBNs for all virus species, partly because of the past opposition of many animal virologists. These virologists, incidentally, dismissed the fact that the eminent animal virologist Frank Fenner[26] had been the first person to use the system. As discussed in Section 8, it is not always clear what sort of democratic process would satisfy the ICTV, or for that matter its critics, and it can only be hoped that virologists will be more inclined in the future to engage in taxonomic debates than they did in the past.[24] Few virologists express an opinion on taxonomic issues with the result that minority views expressed by a few vocal individuals are often heard disproportionally.

The latest official ICTV definition of virus taxon, which masquerades as a definition of virus species and does not accept that classes are indisputable constituents of any classification scheme, testifies to the need for virologists not to follow the desperate call that they should leave taxonomy alone.[93] Frederick Murphy, a Life Member and past President of ICTV, in his contribution to the ICTV forum on the pros and cons of the new ICTV species definition,[64] suggested that a one-day international meeting should be convened to hammer out controversial taxonomic issues that cannot be resolved in the few minutes usually available during a Virology Congress. The present review provides ample material that could be discussed at such a meeting.

References

1. Fauquet CM. Taxonomy, classification and nomenclature of viruses. In: Mahy BWJ, Van Regenmortel MHV, editors. *Desk encyclopedia of general virology*. San Diego: Elsevier; 2010. p. 80–95.
2. Bos L. The naming of viruses: an urgent call to order. *Arch Virol* 1999;**144**:631–6.
3. Van Regenmortel MHV. Viruses are real, virus species are man-made taxonomic constructions. *Arch Virol* 2003;**148**:2483–90.
4. Quine WV. *Word and object*. Cambridge: MIT Press; 1960.
5. Pringle CR. The 20th meeting of the executive committee of the ICTV. Virus species, higher taxa, a universal database and other matters. *Arch Virol* 1991;**119**:303–4.
6. Buck RC, Hull DL. The logical structure of the Linnaean hierarchy. *Syst Zool* 1966;**15**: 97–111.
7. Mahner M, Bunge M. *Foundations of biophilosophy*. Berlin: Springer-Verlag; 1997.
8. Gibbs AJ, Armstrong JS, Gibbs MJ. A type of nucleotide motif that distinguishes tobamovirus species more efficiently than nucleotide signatures. *Arch Virol* 2004;**149**: 1941–54.
9. Van Regenmortel MHV. Virus species. In: Tibayrenc M, editor. *Genetics and evolution of infectious diseases*. London, Burlington: Elsevier; 2011. p. 3–19.
10. Quine WV. Classes versus properties. In: *Quiddities: an intermittently philosophical dictionary*. London: Penguin Books; 1990. p. 22–6.
11. Bernier R. The species as an individual: facing essentialism. *Syst Zool* 1984;**33**:460–9.
12. Ghiselin MT. A radical solution to the species problem. *Syst Zool* 1974;**23**:536–44.

Downloaded From Vet-eBooks.com

13. Hull DL. Are species really individuals? *Syst Zool* 1976;**25**:174−91.
14. Hull DL. *Science as a process*. Chicago: The University of Chicago Press; 1988.
15. De Queiroz K, Donoghue MJ. Phylogenetic systematics and the species problem. *Cladistics* 1988;**4**:317−38.
16. Ruse M. Biological species: natural kinds, individuals, or what? *Brit J Philos Sci* 1987;**38**: 225−42.
17. Van Regenmortel MHV. The metaphor that viruses are living is alive and well, but it is no more than a metaphor. *Stud Hist Philos Biol Biomed Sci* 2016. http://dx.doi.org/10.1016/j.shpsc.2016.02.017.
18. Ereshefsky M. Darwin's solution to the species problem. *Synthese* 2009. http://dx.doi.org/10.1007/s11229-009-9538-4.
19. Claridge MF. Species are real biological entities. In: Ayala FJ, Arp R, editors. *Contemporary debates in philosophy of biology*. Chichester, UK: Wiley-Blackwell; 2010. p. 91−109.
20. Mishler BD. Species are not uniquely real biological entities. In: Ayala FJ, Arp R, editors. *Contemporary debates in philosophy of biology*. Chichester, UK: Wiley-Blackwell; 2010. p. 110−22.
21. Mayr E. *Populations, species and evolution*. Cambridge, Mass: Harvard University Press; 1970.
22. Mayr E. *The growth of biological thought. Diversity, evolution and inheritance*. Cambridge, MA: Harvard University Press; 1982.
23. Milne RG. The species problem in plant virology. *Microbiol Sci* 1984;**1**:113−22.
24. Matthews REF. The history of viral taxonomy. In: Matthews REF, editor. *A critical appraisal of viral taxonomy*. Boca Raton, Florida: CRC Press; 1983. p. 1−35.
25. Wildy P. Classification and nomenclature of first report of the international committee on nomenclature of viruses. *Monogr Virol* 1971;**5** [Basel (Karger)].
26. Fenner F. The classification and nomenclature of viruses. Second report of the international committee on taxonomy of viruses. *Intervirology* 1976;**7**:1−115.
27. Francki RIB, Fauquet CM, Knudson DL, Brown F. Fifth report of the international committee on taxonomy of viruses. *Arch Virol Suppl* 1991;**2**:450.
28. Gibbs AJ, Harrison B. *Plant virology. The principles*. London: Edward Arnold; 1976.
29. Kingsbury DW. Species classification problems in virus taxonomy. *Intervirology* 1985;**24**: 62−70.
30. Van Regenmortel MHV. Applying the species concept to plant viruses. *Arch Virol* 1989; **104**:1−17.
31. Colwell RK. Niche: a bifurcation in the conceptual lineage of the term. In: Keller EF, Lloyd EA, editors. *Keywords in evolutionary biology*. Cambridge, Mass: Harvard University Press; 1992. p. 241−8.
32. Beckner M. *The biological way of thought*. New York: Columbia University Press; 1959.
33. Van Regenmortel MHV. Logical puzzles and scientific controversies: the nature of species, viruses and living organisms. *Syst Appl Microbiol* 2010;**33**:1−6.
34. Ball LA. The universal taxonomy of viruses in theory and practice. In: Fauquet CM, et al., editors. *Virus taxonomy: eight report of the international committee on taxonomy of viruses*. Amsterdam: Elsevier Academic Press; 2005. p. 3−8.
35. Van Regenmortel MHV, Bishop DHL, Fauquet CM, Mayo MA, Maniloff J, Calisher CH. Guidelines to the demarcation of virus species. *Arch Virol* 1997;**142**: 1505−18.
36. Van Regenmortel MHV. Introduction to the species concept. In: Van Regenmortel MHV, Fauquet CM, Bishop DHL, Carstens EB, Estes MK, Lemon SM, et al., editors. *Virus*

Downloaded From Vet-eBooks.com

taxonomy − 7th report of the international committee on taxonomy of viruses. New York & San Diego: Academic Press; 2000. p. 3−16.
37. Van Regenmortel MHV. Virus species and virus identification: past and current controversies. *Inf Gen Evol* 2007;**7**:133−44.
38. Van Regenmortel MHV. Virus species. In: Mahy BWJ, Van Regenmortel MHV, editors. *Desk encyclopedia of general virology*. San Diego: Elsevier; 2010. p. 37−41.
39. Van Regenmortel MHV, Mahy BWJ. Emerging issues in viral taxonomy. *Emerg Infect Dis* 2004;**10**:8−13.
40. Gibbs AJ, Gibbs MJ. A broader definition of the 'virus species'. *Arch Virol* 2006;**151**: 1419−22.
41. Van Regenmortel MHV, Ackermann H-W, Calisher CH, Dietzgen RG, Horzinek M, Keil GM, et al. Virus species polemics: 14 senior virologists oppose a proposed change to the ICTV definition of virus species. *Arch Virol* 2013;**158**:1115−9.
42. Fristrup K. Character: current usages. In: Keller EF, Lloyd EA, editors. *Keywords in evolutionary biology*. Cambridge, MA: Harvard University Press; 1992. p. 45−51.
43. Inglis WG. Characters: the central mystery of taxonomy and systematics. *Biol J Linn Soc* 1991;**44**:121−39.
44. Van Regenmortel MHV. Virologists, taxonomy and the demands of logic. *Arch Virol* 2006;**159**:1251−5.
45. Pigliucci M. Species as family resemblance concepts: the (dis-) solution of the species problem? *BioEssays* 2003;**25**:596−602.
46. Van Regenmortel MHV. Virus species: a much overlooked but essential concept in virus classification. *Intervirology* 1990;**31**:241−54.
47. Dekker EL, Dore I, Porta C, Van Regenmortel MHV. Conformational specificity of monoclonal antibodies used in the diagnosis of tomato mosaic virus. *Arch Virol* 1997;**95**: 191−203.
48. Hebert PDN, Gregory TR. The promise of DNA barcoding for taxonomy. *Syst Biol* 2005; **54**:852−9.
49. Ebach MC, Holdrege C. More taxonomy, not DNA barcoding. *Bioscience* 2005;**55**: 822−3.
50. Gorbalenya AE. Phylogeny of viruses. In: Mahy BWJ, Van Regenmortel MHV, editors. *Desk encyclopedia of general virology*. San Diego: Elsevier; 2010. p. 66−9.
51. Villarreal LP. Evolution of viruses. In: Mahy BWJ, Van Regenmortel MHV, editors. *Desk encyclopedia of general virology*. San Diego: Elsevier; 2010. p. 70−80.
52. Calisher CH, Horzinek M, Mayo MA, Ackermann HW, Maniloff J. Sequence analyses and a unifying system of virus taxonomy: consensus via consent. *Arch Virol* 1995;**140**: 2093−9.
53. Lewontin RC. Genotype and phenotype. In: Fox Keller E, Lloyd EA, editors. *Keywords in evolutionary biology*. Cambridge: Harvard University Press; 1992. p. 137−44.
54. Wolf U. The genetic contribution to the phenotype. *Hum Gen* 1995;**95**:127−48.
55. Neumann-Held EM. Let's talk about genes: the process molecular gene concept and its context. In: Oyama S, Griffiths P, Gray R, editors. *Cycles of contigency − developmental systems and evolution*. Cambridge: MIT Press; 2001. p. 69−84.
56. Ponomarenko JV, Van Regenmortel MHV. B cell epitope prediction. In: Gu J, Bourne PE, editors. *Structural bioinformatics*. 2nd ed. John Wiley & Sons Inc.; 2009. p. 849−79.
57. Moss L. Deconstructing the gene and reconstructing molecular developmental systems. In: Oyama S, Griffiths P, Gray R, editors. *Cycles of contigency − developmental systems and evolution*. Cambridge: MIT Press; 2001. p. 85−97.
58. Moss L. *What genes can't do*. Cambridge, Massachusetts: MIT Press; 2003.

Downloaded From Vet-eBooks.com

59. Krupovic M, Bamford DH. Order to the viral universe. *J Virol* 2010;**84**:12476–9.
60. Krupovic M, Bamford DH. Double-stranded DNA viruses: 20 families and only five different architectural principles for virus assembly. *Curr Opin Virol* 2011;**1**:118–24.
61. Van Regenmortel MHV. Biological complexity arises from the ashes of genetic determinism. *J Mol Recognit* 2004;**17**:145–8.
62. Lentz TL. Viruses as ligands of eukaryotic cell surface molecules. In: Gibbs AJ, Calisher CH, Garcia-Arenal F, editors. *Molecular basis of virus evolution*. Cambridge University Press; 1995. p. 135–49.
63. King A, Adams M, Lefkowitz E, Carstens E. ICTV proposal 2011.002sg. In: *2011.002a-uG.A.v9.statute_and_code_changes.pdf*; 2012. p. 15. http://talk.ictvonline.org/files/ictv_official_taxonomy_updates_since_the_8th_report/m/general-2008/4444.aspx.
64. King A. *Comments to proposed modification to code rule 3.21 (defining virus species)*. ICTV Discussions; 2012. http://talk.ictvonline.org/discussions/ictv1/f/63/t/3930.aspx/.
65. Quine WV, Ullian JS. *The Web of belief*. New York: McGraw-Hill; 1978. p. 44.
66. Baldwin JT, Lessmann O. *What is Russell's paradox?* Scientific American; August 17, 1998. http://www.scientificamerican.com/article/what-is-russells-paradox/.
67. Fauquet CM, Bisaro DM, Briddon RW, Brown JK, Harrison BD, Rybicki EP, et al. Virology division news: revision of taxonomic criteria for species demarcation in the family *Geminiviridae*, and an updated list of begomovirus species. *Arch Virol* 2003;**148**:405–21.
68. Bao Y, Kapustin Y, Tatusova T. Virus classification by pairwise sequence comparison (PASC). In: Van Regenmortel MHV, Mahy B, editors. *Desk encyclopedia of general virology*. Oxford: Academic Press, Elsevier; 2010. p. 95–100.
69. Fauquet CM, Briddo RW, Brown JK, Moriones E, Stanley J, Zerbini M, et al. Geminivirus strain demarcation and nomenclature. *Arch Virol* 2008;**153**:783–821.
70. Fauquet CM, Stanley J. Revising the way we conceive and name viruses below the species level: a review of geminivirus taxonomy calls for new standardized isolate descriptors. *Arch Virol* 2005;**150**:2151–79.
71. Padidam M, Sawyer S, Fauquet CM. Possible emergence of new geminiviruses by frequent recombination. *Virology* 1999;**265**:218–25.
72. Adams MJ, Lefkowitz EJ, King AMQ, Carstens EB. Recently agreed changes to the international code of virus classification and nomenclature. *Arch Virol* 2013;**158**:2633–9.
73. Van Regenmortel MHV. Perspectives on binomial names of virus species. *Arch Virol* 2001;**146**:1637–40.
74. Anonymous [International Commission on Zoological Nomenclature]. *International code of zoological nomenclature*. 4th ed. London: International Trust for Zoological Nomenclature; 1999. http://www.iczn.org/.
75. Lapage SP, Sneath PHA, Lessel EF, Skerman VBD, Seeliger HPR, Clark WA, editors. *International code of nomenclature of bacteria*. Washington, DC: ASM Press; 1992. ISBN: 10:1-55581-039-X.
76. McNeill J, Barrie FR, Buck WR, Demoulin V, Greuter W, Hawksworth DL, et al., editors. *International code of nomenclature for Algae, Fungi, and plants (Melbourne code)*. Koeltz Scientific Books; 2012. 232 p. http://www.iapt-taxon.org/nomen/main.php/.
77. King A, Lefkowitz E, Adams MJ, Carstens E, editors. *Virus taxonomy. Ninth report of the international committee on taxonomy of viruses*. Elsevier Academic Press; 2012. 1327 p. ISBN 978-0-12-384684-6.
78. Mayo MA, Van Regenmortel MHV. Ictv and the virology division news. *Arch Virol* 2000;**145**:1985–8.

Downloaded From Vet-eBooks.com

79. Kuhn JH, Dürrwald R, Bao Y, Briese T, Carbone K, Clawson AN, et al. Taxonomic reorganization of the family *Bornaviridae*. *Arch Virol* 2015;**160**:621−32.
80. Vetten HJ, Haenni A-L. Taxon-specific suffixes for vernacular. *Arch Virol* 2006;**151**: 1249−50.
81. Matthews REF. *Plant virology*. San Diego: Academic Press; 1971.
82. Brunt A, Crabtree K, Gibbs A. *Viruses of tropical plants*. Wallingford: CAB International; 1990.
83. Albouy J, Devergne JC. *Maladies à virus des plantes ornementales*. Paris: Éditions de l'INRA; 1998.
84. Bos L. *Plant viruses, unique and intriging pathogens − a textbook of plant virology*. Leiden: Backhuys Publishers; 1999.
85. Matthews REF. Classification and nomenclature. Third report of the international committee on taxonomy of viruses. *Intervirology* 1979;**7**:1−115.
86. Matthews REF. Classification and nomenclature of viruses. Fourth report of the international committee on taxonomy of viruses. *Intervirology* 1982;**17**:1−200.
87. Murphy FA, Fauquet CM, Bishop DHL, Ghabrial SA, Jarvis AW, Martelli GP, et al., editors. *Virus taxonomy − sixth report of the international committee on taxonomy of viruses*. Vienna: Springer; 1995. 586 p.
88. Bos L. Structure and typography of virus names. *Arch Virol* 2000;**145**:429−32.
89. Agut H. Back to Latin and tradition: a proposal for an official nomenclature of virus species. *Arch Virol* 2002;**147**:1465−70.
90. Van Regenmortel MHV, Fauquet CM. Only italicized species names of viruses have a taxonomic meaning. *Arch Virol* 2002;**147**:2247−50.
91. Mayo MA, Horzinek MC. A revised version of the international code of virus classification and nomenclature. *Arch Virol* 1998;**143**:1645−54.
92. Drebot MA, Henchal E, Hjelle B, Le Duc JW, Repik PM, Roehrig JT, et al. Improved clarity of meaning from the use of both formal species names and common (vernacular) virus names in virological literature. *Arch Virol* 2002;**147**:2465−71.
93. Calisher CH, Mahy BMJ. Taxonomy: get it right or leave it alone. *Am J Trop Med Hyg* 2003;**68**:505−6.
94. Kuhn JH, Jahrling PB. Clarification and guidance on the proper usage of virus and virus species names. *Arch Virol* 2010;**145**:445−53.
95. Calisher CH, Van Regenmortel MHV. Should all other biologists follow the lead of virologists and stop italicizing the names of living organisms? A proposal. *Zootaxa* 2009; **2113**:63−8.
96. Bos L. International naming of viruses. A digest of recent developments. *Arch Virol* 2002; **147**:1471−7.
97. Mayo MA. ICTV at the Paris ICV: results of the plenary session and the binomial ballot. *Arch Virol* 2002;**147**:2254−60.
98. Ball LA, Mayo MA. Report from the 33rd meeting of the ICTV executive committee. *Arch Virol* 2004;**149**:1259−63.
99. Van Regenmortel MHV, Burke DS, Calisher CH, Dietzgen RG, Fauquet CM, Ghabrial SA, et al. A proposal to change existing virus species names to non-latinized binomials. *Arch Virol* 2010;**155**:1909−19.
100. Bukreyev AA, Chandran K, Dolnik O, Dye JM, Ebihara H, Leroy EM, et al. Discussions and decisions of the 2012−2014 international committee on taxonomy of viruses (ICTV) Filoviridae Study group, January 2012−June 2013. *Arch Virol* 2014;**159**:821−30.
101. Gibbs AJ. Virus nomenclature descending into chaos. *Arch Virol* 2000;**145**:1505−7.
102. Gibbs AJ. Virus nomenclature: where next. *Arch Virol* 2003;**148**:1645−53.

Downloaded From Vet-eBooks.com

103. Blachowicz J. How science textbooks treat scientific method: a philosopher's perspective. *Brit J Philos Sci* 2009;**60**:303−44.
104. Nola R, Sankey H. *Theories of scientific method*. Stocksfield: Acumen Publishers; 2007.
105. Greuter W, Garrity G, Hawksworth DL, Jahn R, Kirkn PM, Knapp S, et al. Draft BioCode (2011). Principles and rules regulating the naming of organisms. New draft, revised in November 2010. *Bionomina* 2011;**3**:26−44.
106. Van Regenmortel MHV, Mayo MA, Fauquet CM, Maniloff J. Virus nomenclature: consensus versus chaos. *Arch Virol* 2000;**145**:2227−32.
107. Radoshitzky SR, Bào Y, Buchmeier MJ, Charrel RN, Clawson AN, Clegg CS, et al. Past, present, and future of arenavirus taxonomy. *Arch Virol* 2015;**160**:1851−74.
108. Anonymous [International Committee on Taxonomy of Viruses]. *The universal virus database of the international committee on taxonomy of viruses*. 2002. http://web.archive.org/web/20070611143548/http://phene.cpmc.columbia.edu/index.htm.
109. Buchen-Osmond C, Blaine L, Horzinek MC. The universal virus database of ICTV (ICTVdB). In: Van Regenmortel MHV, Fauquet CM, Bishop DHL, Carstens EB, Estes MK, Lemon SM, et al., editors. *The seventh report of the international committee on taxonomy of viruses*. Academic Press, Elsevier Science & Technology Books; 2000. p. 19−24.
110. Van Rijsbergen K. *Information retrieval*. 2nd ed. London: Butterworths; 1979.

Downloaded From Vet-eBooks.com

This page intentionally left blank

Downloaded From Vet-eBooks.com

A Theory-Based Pragmatism for Discovering and Classifying Newly Divergent Species of Bacterial Pathogens

2

F.M. Cohan [1], *Sarah Kopac* [2]
[1]Wesleyan University, Middletown, CT, United States; [2]University of Rhode Island, Kingston, RI, United States

1. Introduction

Are bacterial species real? They are real enough to the systematists who classify them, and to the practitioners of microbiology who depend on bacterial classification. Bacterial systematists have routinely identified species as closely related groups that differ in their disease-causing properties, in their ecological roles in biological communities, and in their physiological capacities.[1] To provide this service, systematists have taken a simple and pragmatic approach—to define species as groups (or clusters) of close relatives separated by large gaps in phenotypic and molecular characters.[1,2] This practical approach has the cachet of approval from no less an evolutionary biologist than Charles Darwin.[3,4] Darwin proposed that animal and plant species should be defined as closely related groups that can coexist as phenotypically distinct clusters,[3–5] and this is largely the approach taken by bacterial systematists. This cluster-based approach has proved to be remarkably robust, even as the criteria for defining bacterial species have changed over the decades, from being based on phenotype (usually metabolism) to whole-genome similarity (as measured through genome hybridization) to sequence identity.[1] Bacterial systematists have argued about whether the species they recognize are too narrowly or too broadly defined, and whether they are using the best criteria for demarcating species, but they have agreed that species should hold the essential property of being clusters of close relatives with gaps between them.[6]

However, many microbiologists and most systematists outside of microbiology have understood species to be more than closely related groups separated by gaps.[7,8] They have viewed the species level of taxonomy as having a reality beyond human attempts at classification. Largely under the influence of Ernst Mayr, the property of cohesion has become understood as a quintessential aspect of species.[7–10] In this view, species are real because they are the largest groups whose diversity is constrained by a force of cohesion. In the case of the highly sexual animals and plants, the force constraining diversity within species is understood to be genetic exchange. In Mayr's biological species concept, speciation requires certain unusual circumstances that allow newly divergent populations to break free of cohesion by recurrent, high-

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00002-0
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

frequency genetic exchange; speciation is therefore understood to be rare.[9] Zoologists have questioned whether animal species are really cohesive across their geographical ranges, and whether cohesion by genetic exchange actually prevents speciation.[11] This controversy has raised our doubts as to whether bacterial species are cohesive,[12] an issue to which we shall return.

Many concepts of species have been developed since Mayr's biological species concept, and most have in common certain quintessential features, most related to cohesion[7,12]: the diversity within a species is limited by a force of cohesion, different species are ecologically distinct and irreversibly separate, and species are invented only once. In what we might consider Mayrian concepts of species, these essential properties have been extended to other groups where genetic exchange is rare or absent, such as the bacteria.[10,13]

With our colleagues, we have developed the "ecotype" theory of bacterial species, where ecotypes are the most newly divergent populations that are ecologically distinct from one another and are each ecologically homogeneous.[14] Different ecotypes can coexist indefinitely as a result of their ecological distinctness. The ecotype theory recognizes that every population will have *some* ecological diversity: ecotypes are defined such that the ecological divergence among lineages within an ecotype is not sufficient to allow them to coexist indefinitely; we refer to ecologically distinct lineages within an ecotype as "ephemeral ecotypes" (Fig. 2.1). We consider the splitting of one ecotype into two to be the fundamental diversity-creating process of speciation in bacteria.[12–17] We have developed various models of speciation within the ecotype concept, some of which demand cohesion while others do not (Fig. 2.2).

Models of species cohesion depend on homogeneity among members of a species. In the case of animals and plants, cohesion across populations by genetic exchange is widely thought to require homogeneity of reproductive features, such that genes can be

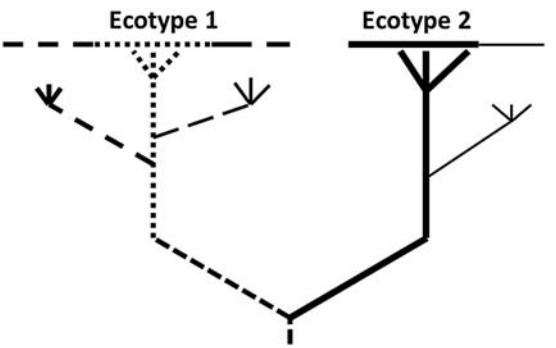


Figure 2.1 Ecological divergence between ecotypes and ecological homogeneity within ecotypes. The ecological divergence between ecotypes is sufficient for them to coexist into the indefinite future. Ecotypes are defined so that any ecological differences among lineages within ecotypes are not sufficient to allow them to coexist indefinitely. We thus refer to ecologically distinct lineages within ecotypes as "ephemeral ecotypes." The different styles of dashed lines within Ecotype 1 refer to different ephemeral ecotypes; note that only one of these lineages persists to the present. The different weights of solid lines represent different ephemeral ecotypes within Ecotype 2.

Downloaded From Vet-eBooks.com

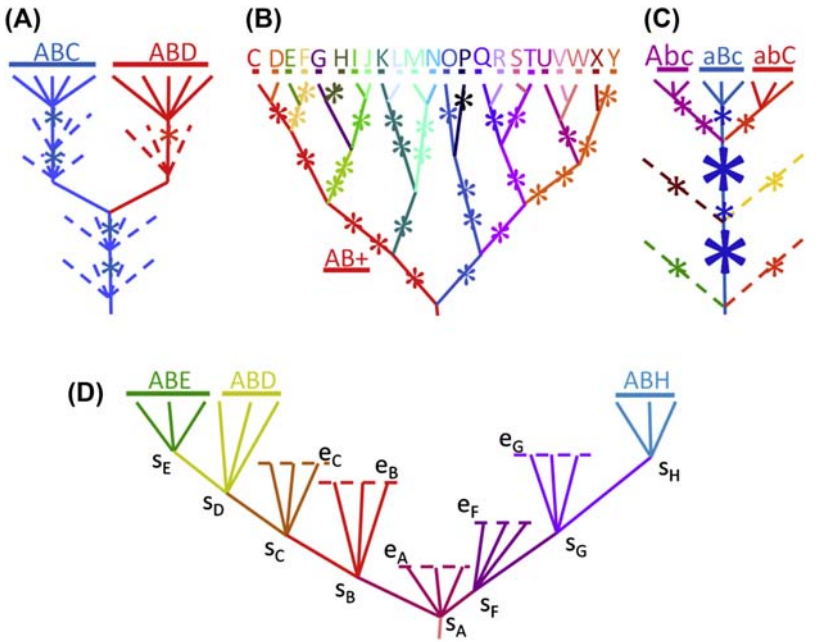


Figure 2.2 Models of bacterial speciation. Ecotypes are represented by different shades of gray; periodic selection events are indicated by asterisks, and extinct lineages are represented by dashed lines. The letters at the top represent the resources that each group of organisms can utilize. In cases where ecotypes utilize the same set of resources but in different proportions, the predominant resource of each ecotype is noted by a capital letter. (A) The Stable Ecotype model. In the Stable Ecotype model, each ecotype endures many periodic selection events during its long lifetime. The Stable Ecotype model generally yields a one-to-one correspondence between ecotypes and sequence clusters because ecotypes are formed at a low rate. The ecotypes are able to coexist indefinitely because each has a resource not shared with the others. (B) The Speedy Speciation model. This model is much like the Stable Ecotype model, except that speciation occurs so rapidly that most newly divergent ecotypes cannot be detected as sequence clusters in multilocus analyses. (C) The Nano-Niche model of bacterial speciation. In the figure, there are three Nano-Niche ecotypes that use the same set of resources but in different proportions (noted by Abc, aBc, and abC). Each Nano-Niche ecotype can coexist with the other two because they have partitioned their resources, at least quantitatively. However, because the ecotypes share all their resources, each is vulnerable to a possible speciation-quashing mutation that may arise in the other ecotypes. (D) The Species-Less model. Here the diversity within an ecotype is not limited by periodic selection but instead by the short time from the ecotype's invention as a single mutant until its extinction. The origination and extinction of each ecotype i is indicated by s_i and e_i, respectively. In the absence of periodic selection, each extant ecotype that has given rise to another ecotype is a paraphyletic group, and each recent ecotype that has not yet given rise to another ecotype is monophyletic.[14] Used with permission from the American Society for Microbiology.

Downloaded From Vet-eBooks.com

successfully exchanged.[9,10] Likewise, the ecotype model is premised on the existence of ecotypes whose members are *ecologically* homogeneous and interchangeable. In some versions of the ecotype model (e.g., the Stable Ecotype model), the ecological homogeneity within an ecotype leads to cohesion. Here, natural selection favoring one competitively superior adaptive mutation within an ecotype causes the mutation to reach 100% frequency within the ecotype. Because recombination is so rare in bacteria,[18] the entire genome of the adaptive mutant can reach nearly 100% frequency.[13,19] Thus, the ecological homogeneity within an ecotype can result in recurrent, genome-wide purges of diversity called periodic selection (Fig. 2.3). Periodic selection is the principal force of cohesion within asexual or rarely sexual lineages of bacteria,[10,13] although as we shall see, cohesion might not be a universal property of bacterial ecotypes.

There is an important pragmatic reason for bacterial species to be demarcated as ecologically homogeneous units. The animal ecologist G. Evelyn Hutchinson saw species as groups that should be homogeneous in their physiological, biochemical, morphological, and ecological characteristics.[20] He noted that species so defined

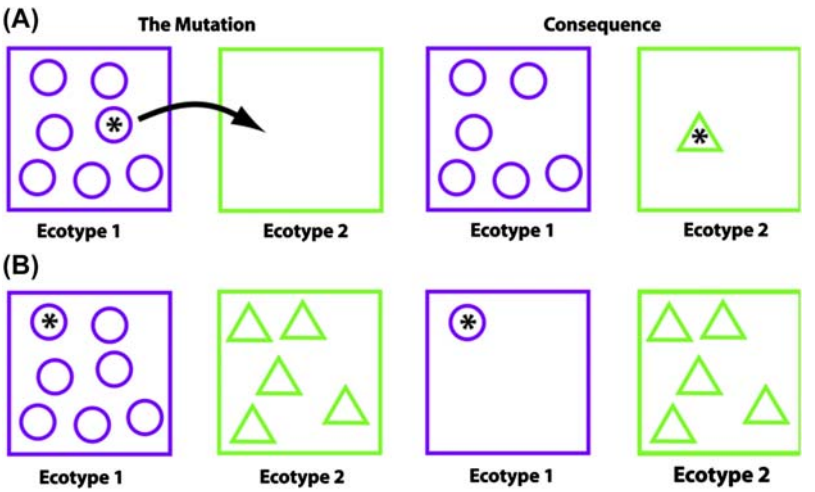


Figure 2.3 The dynamics of ecotype formation and periodic selection within an ecotype. Circles represent different genotypes, and asterisks represent adaptive mutations. (A) Ecotype formation event. A mutation or a recombination event allows the cell to occupy a new ecological niche, founding a new ecotype. A new ecotype can be formed only if the founding organism has undergone a fitness trade-off, whereby it cannot compete successfully with the parental ecotype in the old niche. (B) Periodic selection event. A periodic selection mutation improves the fitness of an individual such that the mutant and its descendants outcompete all other cells within the ecotype; these mutations do not affect the diversity within other ecotypes because ecological differences between ecotypes prevent direct competition. Periodic selection leads to the distinctness of ecotypes by purging the divergence within, but not between ecotypes.[90]
Used with permission from Landes Publishers.

Downloaded From Vet-eBooks.com

have the useful property that the characteristics of any individual classified to a species could be easily predicted. While we believe Hutchinson was overly optimistic about the homogeneity of animal and plant species, microbiologists could probably agree that a taxonomy based on homogeneity, if possible, would be extremely beneficial.[16] For example, under this approach, the total membership of a pathogenic species would have the same disease-causing properties, the same tissue tropisms, the same transmission properties, and the same host range, while organisms with significantly different properties would be recognized as different species.

It is widely understood that the species recognized by bacterial systematics are far from satisfying the Hutchinsonian property of homogeneity. The named species have long been known to be metabolically and ecologically diverse.[17,21–29] There are reasons to suspect that homogeneous, Hutchinsonian species of bacteria may be limited in their phylogenetic breadth; in the extreme, they may even be as phylogenetically narrow as a single cell and its immediate descendants.[30] Genome comparisons suggest the possibility that, at least in some taxa, extremely close relatives are already distinct in their genome content.[14,31–39] That is, bacteria may acquire genes by horizontal genetic transfer at such a high rate that the set of ecologically homogeneous organisms may be too small to be worth the trouble to recognize as a species entity.

The goal of this chapter is to lay out a protocol for determining the phylogenetic extent of ecological homogeneity. Approaches to discovering homogeneous, cohesive species of bacteria are handicapped by various features of bacterial ecology and evolution, which make it difficult to recognize the ecological dimensions by which species diverge or the physiological adaptations that underlie the ecological divergence of new species.[16] This is because we cannot just look at bacteria and infer how they are different ecologically, as we can with closely related birds of different beak size or shape. Also, horizontal genetic transfer is thought to be responsible for the formation of new ecotypes,[40–43] and we cannot predict the genes transferred or their donor source. We therefore cannot always anticipate the dimensions of ecological and physiological divergence among new bacterial species, even in groups that are well characterized.[16]

The discovery of newly divergent bacterial species requires a universal method that is not based on a priori knowledge or intuition about the ecological dimensions of speciation. One approach we outline for discovering the homogeneous species of the bacterial world is ecology blind, where we aim to hypothesize ecotype demarcations from sequence data, confirm the ecological distinctness of ecotypes, and then test for their homogeneity and cohesion, all without a priori knowledge of the ecological dimensions of ecotype distinctness.[40] Another ecology-blind approach is to discover ecotypes as clusters of organisms based on similarity of genome content.[19,36,44] We lay out a process to identify groups of organisms that fit into species that are real, in the sense that they are homogeneous and cohesive; we also allow for the possibility that some (perhaps many) groups of bacteria fit only into reified units of close relatives that are neither homogeneous or cohesive.

Various approaches have been taken to demarcate ecologically distinct groups, and it is becoming apparent that the ecology of a taxon can predict both its rate of speciation and its tendency to be cohesive. Here we review recent findings relating

Downloaded From Vet-eBooks.com

to free-living bacteria as well as pathogens. Finally, we present a new pragmatism for bacterial systematics, which will recognize the real, ecologically homogeneous units of bacterial diversity, where practical, and will recognize reified, heterogeneous units of close relatives where necessary.

2. Ecological Breadth of Recognized Species

The classification scheme of bacterial systematics focuses on finding species that are significantly different from one another in DNA sequence identity, genome content, and physiology,[1] but places almost no emphasis on ensuring that each individual species is homogeneous in any characteristic.[45–47] Under this system, two individuals may be in the same species if they show a critical (previously 97%, now 99%) sequence identity in their 16S rRNA genes.[48] This degree of genetic diversity allows for enormous ecological (and disease-causing) differences within a species, as illustrated by *Escherichia coli*. Members of *E. coli* may be specialized as pathogens or commensals, and may be specialized to colonize the large intestine or other parts of the body.[27,34] These are vastly different environments where the bacteria encounter different extracellular secretions, pH, and notable differences in the extracellular matrix, which they must attach to. Moreover, different *E. coli* populations may be specialized to different hosts[49] and different outside environments.[26] The profound ecological and physiological differences among *E. coli* populations are reflected by huge genomic differences, with some divergent populations sharing fewer than half of their genes.[31] Other named species have also been found to contain a high diversity of ecologically, physiologically, and genomically distinct members.[37,38,50–54]

How did systematists come to agree to house such a huge amalgam of diversity within the species they recognize? In the case of the animals and plants, humans have developed an "umwelt," a foundation for demarcating natural groups of consequence for survival, through natural selection and cultural evolution in humans.[55] However, the bacteria escaped the attention of human interest in biodiversity, and so systematists of bacteria had to develop a way of seeing and classifying the diversity of bacteria from scratch.[12] Moreover, as we have mentioned, bacterial systematists have not had the advantage of being able to anticipate either the ecological differences between close relatives of bacteria or the physiological differences underlying their ecological divergence.

Bacteriologists were successful from the middle of the 20th century in developing an objectively based umwelt for species demarcation. While limited at the time to metabolic and other phenotypic characteristics, "numerical taxonomy" allowed bacterial systematists to develop standard levels of phenotypic diversity within and between species[55,56] (Fig. 2.4). In principle, species could have been defined narrowly, even on metabolic grounds.[46] However, systematists made a pragmatic, but fateful decision early on to include strains within a species that were heterogeneous in the presence versus absence of many metabolic capabilities.[1,12] Bacterial species were from the start defined to be extremely diverse in their physiological and hence ecological characteristics.

Downloaded From Vet-eBooks.com

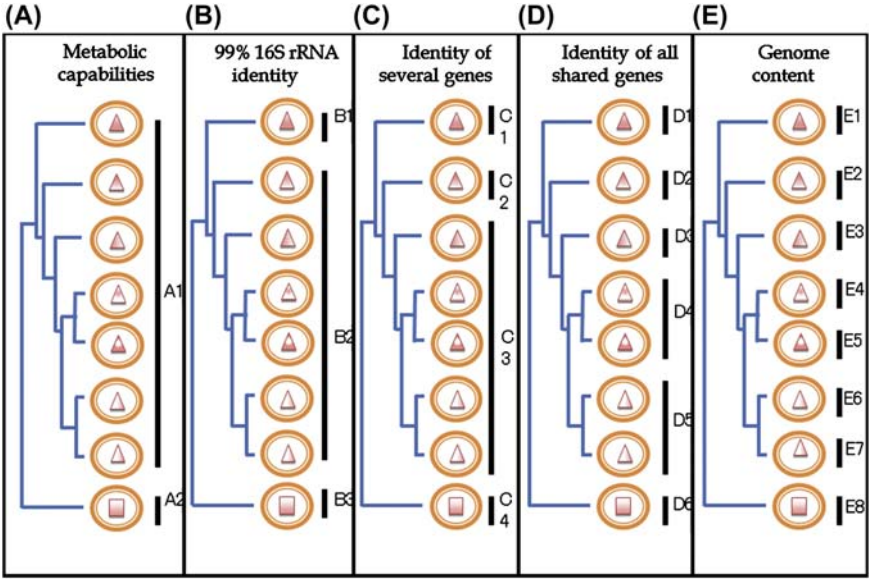


Figure 2.4 Species demarcations under different criteria. Each oval represents a set of closely related cells with identical characteristics of metabolism and ecology, sequences of shared genes, and genome content. Different shapes within the ovals (*triangle* vs. *square*) represent extremely divergent metabolic capabilities (correlated with ecological function), and variations in shading within a particular shape represent more subtle divergence in metabolism and ecology. The species demarcations under each criterion are indicated by a black vertical bar and a species label (e.g., A1). (A) Species were originally defined as groups that differ to a large extent in metabolic capability (indicated by *triangle* vs. *square*), frequently with much metabolic diversity within each species (indicated here by shading differences within the triangles). (B) Defining a species as a group of organisms sharing at least 99% 16S rRNA identity can split the metabolically defined species in the previous panel, as seen here by the splitting of species A1 into B1 and B2. (C) Defining species as clusters based on several protein-coding gene sequences can split a 16S-defined species into groups that are each more ecologically homogeneous. This is seen here by the splitting of species B2 into C2 and C3. (D) Defining species as clusters based on sequence identity for all shared genes can divide species even further, with for example, species C3 being split into D3, D4, and D5. This may be the most highly resolving method for identifying species based on sequences of shared genes. Within species D4, we can see the possibility that even with this level of resolution for species demarcation, there may still be ecological heterogeneity (indicated by the difference in shading between cells in species D4). Species D5 shows an alternative model where this high level of resolution finds clusters that are ecologically homogeneous, as noted by the same shading patterns among members of D5. (E) Defining species by identity of genome content could spuriously split close relatives that are ecologically identical into different species. Note that the two organisms within D5, with the same ecology, are split on the basis of genome content into different species. In this case, E6 and E7 would most likely be different for phage or insertion sequence genes that do not specify ecological niche.

Downloaded From Vet-eBooks.com

Subsequent incorporation of molecular technology has improved species identification in some important ways.[1,12] Molecular approaches have provided universal and readily available methods and criteria for species demarcation to all systematists. Using sequence-based criteria, systematists have been able to avoid recognition of polyphyletic groups. Also, because systematics has been based to some extent on whole genome assays, such as DNA−DNA hybridization, classification has not been deeply affected by recombination across species. Finally, universally applying molecular criteria have led to a pragmatic demarcation scheme that most systematists can agree on[1] (Fig. 2.4).

Nevertheless, molecular technology has not brought about a refinement in the breadth of diversity subsumed within a recognized taxon. Rather, as each new technology has been embraced, including DNA−DNA hybridization,[57] 16S rRNA sequence,[48] multilocus sequence analysis,[6] and genome-wide average nucleotide identity,[58,59] systematists have attempted to calibrate every newest method to yield the existing species taxa.[16,46]

Thus, while the approaches of systematics have brought pragmatic solutions for the practice of systematists, we might ask whether these approaches have been pragmatic for microbiologists outside of systematics. The problem is that when systematists reify an amalgam of ecological and functional heterogeneity into a species taxon, other microbiologists tend to assume that each such species constitutes a natural and fundamental unit of biodiversity. This has led to numerous unfortunate consequences for microbiologists outside of systematics. One such consequence is the classification of genes within a recognized species as essential, "core" genes, shared by all "species" members, versus the nonessential, "dispensable"[39] or "flexible"[54] genes that are shared only by a subset of species members. This dichotomy is false because it is based on the reification of the named species. A more finely demarcated species taxon would yield more core genes and fewer dispensable genes within recognized species. Also, this gives the impression that those genes held only by one subclade are somehow not essential to the ecology or physiology of that group.

This reification of the core genome may have a real, negative impact on vaccine development. Vaccine development can be based on choosing a target protein that is a member of the core genome.[39] However, if the pathogenic strains of concern constitute only a single ecotype within the species diversity, the choice of vaccine target is unnecessarily restricted to the small core genome of the entire named species, rather than the larger set of genes shared among members of the pathogen ecotype.

The broad definition of recognized species has led to innumerable errors in estimation of the critical population genetic parameters of evolution. For example, attributing the name *E. coli* to the huge diversity of ecological specialists within the species gives population geneticists the impression that they are dealing with a group of ecologically interchangeable organisms, such as the members of the fruit fly species *Drosophila pseudoobscura* within a particular habitat. This has led to incorrect application of various algorithms for estimating effective population size and recombination rates,[49] which assume that the organisms sampled are interchangeable.[60] In the case of estimating effective population size from sequence diversity, ecological heterogeneity artificially increases the sequence diversity, and thereby the estimate of population

Downloaded From Vet-eBooks.com

size, by conflating the divergence between populations (which is not affected by population size) with divergence within them. Sequence-based estimations of migration rates have also erred by pooling within a taxon a number of ecologically distinct groups.[61]

In addition to the errors caused by species reification, the broad brush of systematics has also incurred an opportunity cost for different subfields of microbiology, starting with systematics itself. When a systematist discovers a new species and sees that its diversity can be placed within one new species taxon, there is no further motivation from systematics to further explore the ecologically distinct clades within the species. Hence, the research in systematics is impoverished by a standard of detail that leaves much of a clade's diversity uncharacterized.

The broad brush also incurs an opportunity cost on epidemiology. In preparation for the next epidemic, epidemiologists might find it useful to identify all the ecologically distinct populations that already exist within a named pathogenic species. We could then prepare for a future epidemic by characterizing, in advance, the disease-causing properties of each population.[16] The value of this approach is illustrated by our lack of preparedness for the Zika virus epidemic. Based on the dangers presented by the close relatives of Zika, including the yellow fever and dengue viruses, virologists might have been inspired to fully characterize the Zika virus before it brought a worldwide health crisis. This would have led to development of animal models and antibody tests specific to Zika to bear on any future emergence of the virus. In the same vein, we suggest that characterizing the ecotypes closely related to any known pathogen, whether viral or bacterial, can give us a head start on fighting a newly emergent pathogen.

Biotechnologists could also take advantage of a more fine-grained systematics of species. After discovering a strain with a valuable enzyme, one could then search for homologs of the enzyme across closely related, ecologically distinct populations, if they were highlighted by taxonomic recognition.[16,62]

The molecular revolution has taken us far beyond the early days of systematics, when species demarcation was based entirely on metabolism and other phenotypic traits. Sequencing has now revealed ecologically distinct populations within the recognized species, yet we do not take advantage of this information to refine the demarcations of species. This would be an auspicious time to incorporate the high resolution of molecular technology into our taxonomy, so that the physiological and ecological diversity we know to exist within the named species can be officially recognized. An important challenge is to develop universal algorithms to analyze sequence and genome content data to identify populations that are each ecologically homogeneous and ecologically distinct from one another.

3. The Stable Ecotype Model of Bacterial Speciation

In order to integrate ecology into taxonomic classification, we need to take into account the various ways that bacterial species are formed and that diversity within species is constrained. In the Stable Ecotype model, ecotypes are long lived, different

Downloaded From Vet-eBooks.com

ecotypes coexist indefinitely, and speciation is slow (i.e., new ecotypes are formed infrequently). The long-term coexistence of different ecotypes may be fostered by a qualitative ecological divergence, where each ecotype utilizes some unique resource not shared with others[14] (Fig. 2.2). The longevity of each ecotype provides ample opportunity for the ecotype to acquire a unique set of neutral mutations in each gene in the genome; the longevity also provides opportunities for many periodic selection events to occur during the long lifetime of an ecotype. Thus, in the Stable Ecotype model, each ecotype is cohesive by virtue of periodic selection events that recurrently purge the ecotype of sequence diversity.[16]

Ecotypes are founded when a single individual acquires a mutation (or a recombination event) that changes its ecology, through utilizing a new set of resources, thriving under a new set of environmental conditions, or adopting some other change in lifestyle. Since new ecotypes are each founded by a single individual, they start out with zero diversity. A new ecotype is not in direct competition with the members of the parental ecotype because it lives in a different microhabitat or uses at least somewhat different resources. For example, a member of a primarily impetigo-causing (skin-infecting) ecotype of *Streptococcus pyogenes* might mutate or acquire a gene that allows it to primarily infect the throat,[63] thus founding a new ecotype. Although the new ecotype may share the same host species as the parental ecotype, it is utilizing different host resources, and so the two ecotypes may not experience the same periodic selection events.

In the Stable Ecotype model, ecotypes have all the fundamental characteristics broadly attributed to species.[14] Like species, ecotypes are ecologically distinct. They are cohesive because an ecotype's diversity is recurrently purged genome-wide by periodic selection events. Ecotypes are irreversibly separate because their ecological distinctness prevents periodic selection within one from extinguishing the other and because recombination in bacteria is insufficient to reverse ecological diversification.[19]

One consequence of many recurrent periodic selection events in each of several longstanding ecotypes is that closely related ecotypes are expected to correspond to sequence clusters, for any gene in the genome.[64] This is because, while diversity in each gene in the genome is recurrently purged within an ecotype, different longstanding ecotypes are accumulating unique mutations in every gene and new ecotypes are rare. For example, close relatives of *Mycobacterium tuberculosis* form sequence clusters that are each adapted primarily to a different host species.[23] Because each sequence cluster appears to be ecologically homogeneous, the clusters likely represent different ecotypes.

One may straightforwardly test whether a given taxon is subject to the slow speciation and long-lived ecotypes required by the Stable Ecotype model. If ecological diversification is slow, then DNA sequence clusters should reveal ecologically distinct groups that are each ecologically homogeneous. Moreover, there should be appreciable sequence diversity (for example, as much as 0.5−2.0% divergence) accumulated within each ecologically homogeneous sequence cluster. The first step to testing the Stable Ecotype model is to hypothesize and demarcate putative ecotypes as clusters based on sequence data, using one of several algorithms. The next steps

Downloaded From Vet-eBooks.com

are to test whether the putative ecotypes are ecologically distinct from one another and whether members of the same putative ecotype are ecologically homogeneous.[14]

4. Demarcating Putative Ecotypes From Sequence Data

There are multiple algorithms available for hypothesizing ecotype demarcations from sequence data, including AdaptML,[28] Ecotype Simulation,[17] BAPS,[65] and GYMC.[66] In contrast to the approaches of bacterial systematics, none of these algorithms assumes a universal criterion for demarcation. Rather, each algorithm uses sequence data from the taxon of focus to identify the appropriate sequence divergence criterion for distinguishing ecotypes. For example, Ecotype Simulation identifies sequence clusters that are most consistent with ecotypes, assuming that ecotype formation and periodic selection occur at rates inferred from the sequence data.[67]

AdaptML differs from the other algorithms in requiring the habitat of isolation as input data, while the others are blind to ecology (i.e., no information about the ecology or habitat of the strains is taken into account in the analysis).[28] Both approaches have their advantages.[40] AdaptML is useful when associations with certain habitats are suspected, as this algorithm can simultaneously discover ecotypes and confirm their preferences to habitats specified by the investigator. In contrast, the ecology-blind algorithms do not require the researcher to know anything about the potential environmental differences being analyzed. As a result, multiple ecotypes can be found even in environments that were a priori thought to be homogeneous.

The ecotypes hypothesized by these algorithms have consistently been confirmed to be ecologically distinct, based on differences in their habitat associations.[17,28,45,68] It is important to note that ecotypes need not be *absolutely* specialized to different habitats[28]: closely related ecotypes frequently are only somewhat ecologically specialized, and coexist by quantitative differences in habitat distinctness.[28] We have confirmed the ecological distinctness of putative ecotypes in soil *Bacillus* by differences in their habitat associations, including differences in solar exposure, soil texture, rhizospheres, and elevation[17,29] (Kopac et al., unpublished data). Putative ecotypes of *Synechococcus* from hot spring mats have differed in their temperature and depth associations,[45,68] putative ecotypes of *Legionella* have differed in their host ranges,[24] and putative ecotypes in the marine taxon *Vibrio splendidus* have differed in the sizes of particles they were attached to and in their seasons of abundance.[28] In addition, many ecologists have noted that very closely related sequence clusters (demarcated by eye, rather than by a computer algorithm) were different in their habitat associations.[22,23,26,27,69] To sum up, putative ecotypes across a great diversity of phyla (Firmicutes, Proteobacteria, Spirochetes, Actinobacteria, and Cyanobacteria) that were identified as sequence clusters have all appeared ecologically distinct.

In addition, putative ecotypes may be further confirmed to be ecologically distinct through finding physiological and genomic differences that underlie their habitat associations. For example, putative ecotypes of *Bacillus subtilis* associated with more direct solar exposure were found to have membrane differences yielding greater thermal tolerances.[29] Putative ecotypes of *Synechococcus* farther from the source of the

Downloaded From Vet-eBooks.com

hot spring were found to be less tolerant of extreme temperatures[70] and had genes enabling utilization of ions that are most abundant in the downstream part of the spring.[36,71]

5. Ecological Diversity Within Putative Ecotypes

Sequence-based algorithms are frequently not always successful in identifying putative ecotypes that are ecologically homogeneous. Many organisms sampled from a single putative ecotype have turned out to be ecologically divergent from one another. For example, we surveyed genomic diversity within one putative ecotype of *B. subtilis*,[14,29] and found that each of the five isolates sampled had a unique history of positive selection, indicating that no two members of the putative ecotype were ecologically identical.[14,72] Evidence from positive selection, as well as growth experiments, suggested that ecotypes form at an extremely rapid rate in *Bacillus* and that the Stable Ecotype model does not apply to this taxon.[73]

We found a very different pattern of diversification in the hot spring cyanobacterial genus *Synechococcus*. Here, each putative ecotype identified as a cluster with about 0.5% diversity genome-wide appeared to be ecologically homogeneous. The evidence was that each putative ecotype consisted of various sequence types that maintained the same relative frequencies across a great diversity of natural and experimentally perturbed habitats.[68] Thus, in the time that these putative ecotypes accumulated 0.5% sequence diversity across their genomes, they had not diversified ecologically, and so diversification in *Synechococcus* appears to abide by the Stable Ecotype model.

The contrast between *Bacillus* and *Synechococcus* in their rates of speciation suggested a hypothesis to predict which organisms follow the slow speciation of the Stable Ecotype model. Among free-living bacteria, we have predicted that generalist heterotrophs, such as *Bacillus*, with many options for metabolic diversification, will speciate rapidly.[73] On the other hand, photoautotrophs, such as *Synechococcus*, minimally utilize organic compounds, and so they may be much more limited in their ecological opportunities for speciation.

This hypothesis was supported more generally by a 2016 metagenomic study of diversity at various depths within a lake in Wisconsin, United States.[73,74] In a longitudinal survey, Bendall et al. assembled metagenomic sequences into clusters of genetically similar organisms, usually with up to 2% sequence divergence, and one such cluster was shown to lose its diversity in a genome-wide sweep over a span of 8 years. This constituted the first direct evidence for a periodic selection event in nature. The authors also found evidence of some genome-wide sweeps occurring before the study began. Each cluster of organisms that they found swept of diversity, genome-wide, was interpreted as evidence that the entire cluster was ecologically homogeneous, such that one adaptive mutant could outcompete the entire diversity of the cluster. Because these clusters failed to diversify in the time that they accumulated as much as 2% sequence divergence, we may conclude that diversification within these clusters occurred at a slow rate consistent with the Stable Ecotype model. Interestingly, these clusters could be predicted to have little opportunity to diversify, as each was very

Downloaded From Vet-eBooks.com

limited in the carbon resources it could utilize—these clusters were photoautotrophs of the phylum Chlorobi or heterotrophs of single-carbon molecules of various phyla.[73]

The metagenomic study also found various taxa undergoing less profound sweeps, where only a single chromosome region was swept within a cluster.[74] These clusters were interpreted to contain not a single, ecologically homogeneous population, but rather an amalgam of many newly divergent, ecologically distinct ecotypes,[73] following the Adapt Globally Act Locally model.[19,73,75–77] That is, an adaptive mutation that was generally useful across the diversity of ecotypes within a cluster could cause a genome-wide sweep within one ecotype, then transfer as a small recombinant segment to another ecotype, causing a genome-wide sweep there, and so on. The result is that the entire cluster would be swept of diversity only in the gene region around the adaptive mutation and would be heterogeneous everywhere else on the genome. Hence, the clusters undergoing sweeps over just a small chromosomal region were interpreted as undergoing rapid speciation,[73] where a great diversity of ecotypes could be found within 2% sequence divergence. These taxa were for the most part generalist heterotrophs, such as *Bacillus*, supporting the hypothesis that a highly plastic metabolism is the key to rapid speciation.[73,74]

To sum up the data from free-living organisms, the Stable Ecotype model of slow speciation appears to apply to organisms with little opportunity to diversify through utilization of different organic resources, while more rapid speciation occurs in taxa that utilize a vast diversity of organic resources.[73]

Where do bacterial pathogens fit in between the extremes of metabolic plasticity from *Bacillus* to *Synechococcus*? Some pathogens may follow the slow diversification of the Stable Ecotype model, especially if they diversify primarily by adapting to new host species and new host tissues. That is, if susceptible host species and tissues are not numerous, speciation might be infrequent. For example, the *M. tuberculosis* complex (which includes several recognized, taxonomic species) includes sequence clusters that are each ecologically homogeneous and associated with different clades of hosts (humans, artiodactyls, pinnipeds).[78] Also, within *Anaplasma phagocytophilum*, a tick-borne pathogen, sequence clusters have diversified into associations with different host species and are each ecologically homogeneous.[53]

Another pathogen likely following the Stable Ecotype model is *Borrelia burgdorferi* sensu stricto, the spirochete responsible for Lyme disease in North America. This is a tick-borne disease that is maintained in forests through multiple mammalian hosts and their ticks. A multilocus sequence analysis shows multiple clusters, several of which appear specialized to different rodent species,[79] although it is not yet clear whether the adaptations to specific hosts represent genome-wide adaptations or are due only to a single outer-surface protein.[80]

The Stable Ecotype model may reasonably apply to two closely related pathogens within *Yersinia* that infect different tissues. *Yersinia pseudotuberculosis* is transmitted by the fecal–oral route; however, it has the capacity to be lethal if it should invade the lungs or the blood.[81] *Yersinia pestis*, the plague pathogen, emerged from *Y. pseudotuberculosis* and has developed a lifestyle of systemic infection and transmission by fleas.[82] As each taxon appears to be homogeneous in its ecology, these two

Downloaded From Vet-eBooks.com

taxa may be considered ecotypes that are distinguished by host tissue and mode of transmission, while both are generalists with respect to host species.

For pathogens where speciation events are infrequent (for a lack of either susceptible, novel host species, or possible tissues to infect), we hypothesize that the Stable Ecotype model is likely to apply, with many periodic selection events occurring within the long lifetime of any given ecotype. We next consider alternative models of speciation, where ecological diversification is more rapid than can be accommodated by the Stable Ecotype model.[16]

6. Models of Frequent Speciation

6.1 Speedy Speciation Model

In the Speedy Speciation model, cohesion occurs through periodic selection, just as in the Stable Ecotype model.[16] The difference is that speciation is greatly accelerated as occurs in an adaptive radiation. The practical consequence of the rapid speciation is that there are many newly divergent species that cannot be distinguished by neutral divergence in a small set of randomly sampled genes (i.e., genes not involved in the adaptive divergence between species). Depending on the rate of speciation, species could perhaps be distinguished by neutral sequence variation if instead the whole genome were sequenced. Moreover, sequencing of the whole genome may reveal the genes responsible for ecological divergence.

A likely example of Speedy Speciation comes from a study of diversity of *Pseudomonas aeruginosa* within the lungs of cystic fibrosis (CF) patients.[83] Jorth et al. found that in different regions of a CF lung, multiple populations had diversified in situ and were divergent in numerous features that apparently adapted them to their local environments, including differences in nutrient requirements, antibiotic resistance, and virulence. These ecotypes have certainly evolved quickly (within one short-lived human), and the accumulated adaptive differences over time probably represent periodic selection events within each; so the Speedy Speciation model appears to apply. Alternatively, the situation would be different if *P. aeruginosa* ecotypes adapted to CF lungs were to spread frequently from one CF patient to another.[84] In this case the Speedy Speciation model probably would not apply, as a given lung niche could be colonized by an already-adapted ecotype rather than requiring evolution in situ.

A similar case of Speedy Speciation appears to occur in *M. tuberculosis*. It appears that, owing to bottlenecks in transmission of the tuberculosis pathogen between humans, individual hosts are often infected by a single lineage, which can diversify rapidly within a new host. Thus, a given tuberculosis patient is likely to host a diversity of ecologically divergent ecotypes that have evolved in situ.[85] In both *P. aeruginosa* and *M. tuberculosis*, many ecotypes originating within one human's lungs may not be transmitted and so are short lived.

Some pathogens may diversify very quickly in much the same way as free-living generalist heterotrophs. This is because many pathogens actually have a free-living, generalist heterotroph stage between host infections. We may speculate that a

Downloaded From Vet-eBooks.com

pathogen lineage that is adapted to a particular group of hosts may diversify into distinct ecotypes that specialize on different environmental carbon sources. Pathogens with a free-living stage include many members of the Enterobacteriaceae, such as *E. coli*,[86] and some Firmicutes, such as members of *Listeria*.[87] If a pathogen with a free-living phase evolves adaptations to open-environment resources as quickly as *Bacillus* evolves adaptations to soil resources, the newest ecotypes may not be distinguishable as sequence clusters.

6.2 Species-Less Model

The Species-Less model is profoundly different from the Stable Ecotype model and all models that assume cohesion within species.[16] The Species-Less model assumes both rapid speciation and rapid extinction, leading to a high turnover of species. In this case, a species may not persist long enough, from its time of origin to its extinction, to undergo even a single periodic selection event. In the Species-Less model, each ephemeral ecotype, while ecologically homogeneous, could not be considered a cohesive unit. Like the case of the Speedy Speciation model, where species are cohesive, the Species-Less model will lead to a diversity of ecotypes (in this case, ephemeral ecotypes) that cannot be easily distinguished as sequence clusters.

In the Species-Less model, ecotypes evolve not by becoming more efficient in utilizing their current ecological niche, but instead by evolving to invade a new ecological niche. The Species-Less model may apply to the case of pathogens, where immune-escape mutations may each constitute a new ephemeral ecotype.[12,88] Also, the Species-Less model may apply in cases where an environment undergoes a succession process, where organisms at a site must adapt to rapidly changing conditions, for example, the successions that occur on mine tailings, with pH and oxidation levels changing predictably and quickly.[89]

6.3 Nano-Niche Model

Like the Species-Less model, the Nano-Niche model also assumes a high rate of speciation, but here cohesion occurs over a set of different ephemeral ecotypes.[16] In the Nano-Niche model, closely related, ephemeral ecotypes are subtly and only quantitatively different in their ecology. These "Nano-Niche ecotypes" use the same set of resources and conditions, but they coexist much like closely related animal species by using their shared resources and conditions *in different proportions*. Not having any unique resources that might constitute a haven from competition from other ecotypes, each ecotype is ephemeral and vulnerable to extinction from competition with other ecotypes. For a time, the various Nano-Niche, ephemeral ecotypes may coexist while each has its own private periodic selection events. At some point, however, an extremely competitive adaptive mutant (which we call a speciation-quashing mutation) from one ephemeral ecotype may extinguish not only the other members of its own population, but also other closely related, ephemeral ecotypes.[90] In the Nano-Niche model, divergence among very closely related ephemeral ecotypes is limited by these speciation-quashing mutations, and many ephemeral ecotypes might not last long enough to appear as separate sequence clusters.

Downloaded From Vet-eBooks.com

We found evidence for the Nano-Niche model in *B. subtilis* in desert soils from genome sequencing of five members of one putative ecotype.[14] All these isolates were ecologically distinct (as described earlier), but especially interesting was that the strains showed no differences in genome content that would indicate unshared resources. Thus, the strains appeared to be ecologically distinct only in the extent to which they utilize shared resources. For example, all strains sampled had the capacity to utilize maltose, but one strain had additional genes for maltose utilization and was able to grow faster on maltose than the others. These strains may have a limited future of coexistence and likely constitute ephemeral ecotypes, since a speciation-quashing event originating in one ecotype could extinguish the others.

The Nano-Niche model may also apply to bacterial ecotypes that adapt over a long course of infection within a host individual (e.g., a commensal or chronic pathogen in one person). The course of evolutionary adaptation to one human body may bring about multiple periodic selection events within that population, and different individual humans could support different ecotypes. However, the individual hosts might not be different enough to support indefinite coexistence of individual-specific ecotypes. Any speciation-quashing mutation, which makes an individual bacterium not just superior in its own host but also in other hosts, would put an end to the speciation among the various Nano-Niche ecotypes adapting to different host individuals. In our quest to identify ecologically homogeneous groups, we should perhaps be satisfied with finding *sets* of Nano-Niche ecotypes whose diversity will be purged with a speciation-quashing mutation, rather than identifying every ephemeral ecotype.

7. Other Models Where Ecotypes Are Not Discernible as Sequence Clusters

7.1 Recurrent Niche Invasion Model

In the recurrent niche invasion model, mobile genetic elements, such as plasmids or phage, may determine bacterial niches.[16] For example, in the case of *Rhizobium*, a bacterial lineage may acquire a symbiosis-encoding plasmid that adapts it as an endosymbiotic mutualist for a particular set of legume hosts; the lineage may then lose that plasmid and gain another, which adapts it to another set of legume hosts. Ecotypes in this model are not irreversibly separate, as a lineage can recurrently leave one plasmid-determined ecotype and then join another.

In the case of *Rhizobium leguminosarum*, there are five sequence clusters that can each be infected with various symbiosis plasmids that adapt the bacteria to vetch or clover.[91] While the five bacterial clusters have no known diagnostic features that would explain their ecological coexistence, one possibility is that these populations are distinguished quantitatively in their propensities to be infected by different symbiosis plasmids, a result suggested by a contingency test of association between the clusters and plasmid types (vetch-adapting vs. clover-adapting, a 5×2 Fisher's exact test, $P = .011$, based on data in[91]). So, perhaps one dimension of ecological distinctness among the five clusters is a quantitative difference in their tendencies to infect vetch versus clover.

Downloaded From Vet-eBooks.com

A similar situation arises within a set of longstanding sequence clusters that include the pathogens *Bacillus cereus* (a gut pathogen of mammals), *Bacillus anthracis* (causing anthrax and systemic infection in mammals), and *Bacillus thuringiensis* (an insect pathogen). These taxa are problematic in that they are classified by these phenotypes, which are provided by niche-determining plasmids. Because the plasmids can move between sequence clusters, the species taxa *B. cereus* and *B. thuringiensis* are both polyphyletic.[92] As in the case of *R. leguminosarum*, a contingency test of four plasmid categories versus five clusters yielded a high degree of association ($P \approx 0$, based on data in[92]). Thus, part of what is contributing to the coexistence of the various longstanding sequence clusters appears to be their propensities toward infection by the different niche-determining plasmids.

7.2 Cohesive Recombination Model

The cohesive recombination model provides another mechanism by which ecologically distinct populations will fail to be recognizable as sequence clusters.[16] As analyzed quantitatively by Hanage et al.[93] bacteria with the highest rates of recombination may exchange genes so frequently that ecotypes do not accumulate sequence divergence in niche-neutral genes,[94] and so we will not be able to discern the ecotypes as distinct sequence clusters. We note, however, that the rates of genetic exchange in bacteria are never sufficient to hinder or reverse adaptive divergence in niche-specifying genes, as we have previously discussed.[19,40,95] Thus, while genetic exchange in rapidly recombining bacteria will not prevent ecotype formation, it may prevent our ability to discover ecotypes using niche-neutral sequence diversity.

7.3 Geotype Plus Boeing Model

In some cases, a single ecotype from a given community may fall into several distinct sequence clusters, as seen in the Geotype plus Boeing model.[16] Provided that a given taxon has not dispersed frequently, geographically isolated members of a single ecotype may diverge into different clusters (geotypes), even while remaining ecologically interchangeable.[96] This would yield a different sequence cluster in each geographical region, a common phenomenon in the systematics of all organisms that do not readily disperse.[97] In the case of bacteria, geotypes may be a source of confusion for systematists if geotypes have historically been isolated, but now with modern human transport, their dispersal has been accelerated. In this case (the Geotype plus Boeing model), members of one ecotype isolated from a single site may contain multiple clusters representing the ecotype's various, formerly isolated geotypes from all over the world.

8. Are Bacterial Ecotypes Cohesive?

We began with the issue of the reality of bacterial species, whether there is something biologically unique about the level of species. The concept of cohesion has been argued to provide a key dynamic property of species through the biological world—

Downloaded From Vet-eBooks.com

that diversity within a species is limited by certain forces but that divergence above the species level is not.[7,10] The ecotype concept (and particularly the Stable Ecotype model) assumes cohesion within ecotypes, in that diversity within an ecotype is limited by periodic selection, but that divergence between them is not. This cohesion requires ecological homogeneity within an ecotype.[16]

However, ecological homogeneity is not sufficient to ensure cohesion by forces, such as periodic selection and drift, which act recurrently over the lifetime of an ecotype. As we have seen in the Species-Less model, it is possible that new, ecologically homogeneous populations may not persist long enough to encounter a periodic selection event before it goes extinct. In taxa incurring a high turnover of bacterial species, with rapid invention and extinction of ecotypes, the only force limiting the diversity within an ecotype may be its short lifetime before extinction. We have previously proposed that some, perhaps most, of the bacterial ecotypes within a taxon may not represent species-like cohesive groups, while others may be long-lasting and cohesive, and may even extend over broad geographical areas. We have proposed a phylogenetic test to determine whether newly formed ecotypes are cohesive groups[12] (Fig. 2.5). For the purpose of building a systematics that might aim to identify and classify all the ecological diversity within a taxon, it is probably sufficient to focus on finding the ecologically homogeneous clades, without concern for their cohesiveness.

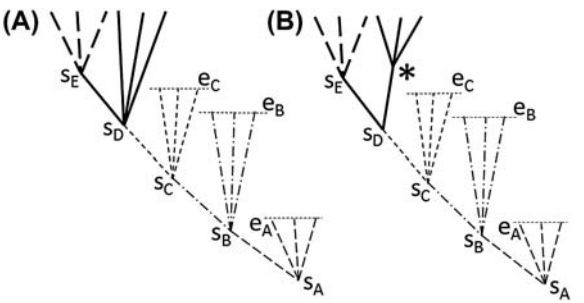


Figure 2.5 The Species-Less model of bacterial diversification. In the Species-Less model, the diversity within an ecotype is not limited by periodic selection but instead by the short time from the ecotype's invention until its extinction. Each ephemeral ecotype in the figure is indicated by a unique line style; the origination and extinction of each ephemeral ecotype i is represented by s_i and e_i, respectively. (A) In the absence of periodic selection, each ephemeral ecotype that has given rise to another ephemeral ecotype is a paraphyletic group (e.g., ecotype D), and each recent ecotype that has not yet given rise to another ecotype is a monophyletic group (e.g., ecotype E). (B) If instead a periodic selection event has occurred in the parental ecotype since it has given rise to a daughter ecotype, the mother−daughter pair will be seen as a pair of sister monophyletic clades. Observing that pairs of most closely related ecotypes are usually observed as paraphyletic−monophyletic pairs would indicate that the origin of new ecotypes is more common than periodic selection in existing ecotypes, giving support to the Species-Less model.[12]
Used with permission from the American Society for Microbiology.

Downloaded From Vet-eBooks.com

9. Incorporating Ecology Into Bacterial Systematics

We suggest that systematics should recognize ecologically homogeneous ecotypes rather than the broadly defined, ecologically heterogeneous amalgams currently recognized. To this end, we lay out a protocol for selecting ecotypes that systematists might have the confidence and motivation to recognize. First, we suggest using sequence data to demarcate ecotypes that appear to represent phylogenetic groups with a history of coexistence as ecologically distinct lineages. Ecotypes could be hypothesized by any of the various universal, sequence-based methods, including AdaptML, Ecotype Simulation, GMYC, and BAPS. If such analyses were to be based on many genes, in the extreme the entire set of shared genes in the genome, more newly divergent ecotypes could be resolved.

Second, the most closely related ecotypes should be confirmed to be ecologically distinct from one another, by differences in habitat association, differences in history of positive selection, or differences in physiology. Physiological differences could be inferred from genomic differences in gene content but ideally would be confirmed experimentally.

Third, in keeping with an important tradition of bacterial systematics, ecotypes should be confirmed to be phenotypically distinct,[1] and we add that ideally the phenotypic differences should confer the ecological niche specificity of the ecotypes.

Fourth, if possible, an ecotype should be confirmed to be ecologically homogeneous, although as we have pointed out, this may be difficult short of sequencing the full genomes of many members of the ecotype.

Fifth, we suggest that we should not be compelled to recognize every ecotype—only those of interest or consequence. This is because some focus taxa may contain multiple, extremely young, ecologically distinct populations that are unlikely to persist into the future (as in the case of the Nano-Niche model). Here we see that there is a conflict between ensuring homogeneity of ecotypes and recognizing only those of potential interest. Thus, the reform we suggest aims to identify the real, ecologically homogeneous groups where possible, but when impractical, we suggest classifying an ecologically heterogeneous clade as an ecotype, provided that it has been identified by sequence-based algorithms as a putative ecotype and has shown to be ecologically distinct from other closely related ecotypes.

Consider first those cases where a recognized, legacy species is found to contain multiple ecotypes, such as is the case for *Bacillus simplex*,[17] *V. splendidus*,[28] and probably many cases where sequence clusters within a pathogenic species are known to differ in host range and/or tissue tropism.[23,26,27,98] In these cases, we suggest keeping the existing species binomial in order to maintain stability of the taxonomy, but suggest adding a trinomial "ecovar" epithet to describe the ecotype taxon. For example, an oak forest−associated and a grassland-associated ecotype within *B. simplex*, from a canyon near Haifa, Israel,[17] might be named *B. simplex* ecovar Alon and *B. simplex* ecovar Esev (based on the Hebrew words for oak and grass). For ecotypes that are found to be outside the phylogenetic range of existing, recognized species, we suggest naming each ecotype as a species.

Downloaded From Vet-eBooks.com

We believe that the approach we have laid out is pragmatic both for systematists and for those whose work would benefit from a full accounting of the ecological diversity among close relatives. The proposed system is pragmatic because it identifies the likely ecotypes through universally available and applicable techniques of genomics and DNA sequencing, as well as computer algorithms to recognize the ecotypes from sequence diversity patterns. And it does not reify heterogeneous groups by attempting to apply a universal molecular criterion to all bacterial species. Microbiologists outside of systematists would benefit from a systematics that would recognize the most recent products of bacterial speciation. Perhaps most importantly, we will more effectively come to know the unique ecological roles played by each member of a vast and diverse microbial community.

Acknowledgments

This work was sponsored by NSF FIBR grant (EF-0328698) and by research funds from Wesleyan University.

References

1. Rosselló-Mora R, Amann R. The species concept for prokaryotes. *FEMS Microbiol Rev* 2001;**25**(1):39−67.
2. Vandamme P, Pot B, Gillis M, de Vos P, Kersters K, Swings J. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol Rev* 1996;**60**(2):407−38.
3. Mallet J. Mayr's view of Darwin: was Darwin wrong about speciation? *Biol J Linn Soc* 2008;**95**:3−16.
4. Darwin C. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London: John Murray; 1859.
5. Stamos DN. *Darwin and the nature of species*. Albany: State University of New York Press; 2007.
6. Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, et al. Opinion: reevaluating prokaryotic species. *Nat Rev Microbiol* 2005;**3**(9):733−9.
7. de Queiroz K. Ernst mayr and the modern concept of species. *Proc Natl Acad Sci USA* 2005; **102**(Suppl. 1):6600−7.
8. Mayr E. *The growth of biological thought: diversity, evolution, and inheritance*. Cambridge: Harvard Univ Press; 1982. p. 295−7 [Chapter 6].
9. Mayr E. *Animal species and evolution*. Cambridge: Belknap Press of Harvard Univ Press; 1963.
10. Templeton A. The meaning of species and speciation: a genetic perspective. In: Otte D, Endler J, editors. *Speciation and its consequences*. Sunderland MA: Sinauer Assoc; 1989. p. 3−27.
11. Mallet J. Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philos Trans R Soc Lond B Biol Sci* 2008;**363**(1506):2971−86.
12. Cohan FM. Are species cohesive? A view from bacteriology. In: Walk S, Feng P, editors. *Bacterial population genetics: a tribute to Thomas S Whittam*. Washington, DC: American Society for Microbiology Press; 2011. p. 43−65.

Downloaded From Vet-eBooks.com

13. Cohan FM. The effects of rare but promiscuous genetic exchange on evolutionary divergence in prokaryotes. *Am Nat* 1994;**143**:965−86.
14. Kopac S, Wang Z, Wiedenbeck J, Sherry J, Wu M, Cohan FM. Genomic heterogeneity and ecological speciation within one subspecies of *Bacillus subtilis*. *Appl Environ Microbiol* 2014;**80**:4842−53.
15. Ward DM. A natural species concept for prokaryotes. *Curr Opin Microbiol* 1998;**1**(3): 271−7.
16. Cohan FM, Perry EB. A systematics for discovering the fundamental units of bacterial diversity. *Curr Biol* 2007;**17**:R373−86.
17. Koeppel A, Perry EB, Sikorski J, Krizanc D, Warner WA, Ward DM, et al. Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. *Proc Natl Acad Sci USA* 2008;**105**:2504−9.
18. Vos M, Didelot X. A comparison of homologous recombination rates in bacteria and archaea. *ISME J* 2009;**3**(2):199−208.
19. Wiedenbeck J, Cohan FM. Origins of bacterial diversity through horizontal gene transfer and adaptation to new ecological niches. *FEMS Microbiol Rev* 2011;**35**:957−76.
20. Hutchinson GE. When are species necessary? In: Lewontin RC, editor. *Population biology and evolution*. Syracuse: Syracuse University Press; 1968. p. 177−86.
21. Barrett EL, Solanes RE, Tang JS, Palleroni NJ. *Pseudomonas fluorescens* biovar V: its resolution into distinct component groups and the relationship of these groups to other *P. fluorescens* biovars, to *P. putida*, and to psychrotrophic pseudomonads associated with food spoilage. *J Gen Microbiol* 1986;**132**(10):2709−21.
22. Manning SD, Motiwala AS, Springman AC, Qi W, Lacher DW, Ouellette LM, et al. Variation in virulence among clades of *Escherichia coli* o157:H7 associated with disease outbreaks. *Proc Natl Acad Sci USA* 2008;**105**(12):4868−73.
23. Smith NH, Gordon SV, de la Rua-Domenech R, Clifton-Hadley RS, Hewinson RG. Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*. *Nat Rev Microbiol* 2006;**4**:670−81.
24. Cohan FM, Koeppel A, Krizanc D. Sequence-based discovery of ecological diversity within *Legionella*. In: Cianciotto NP, Abu Kwaik Y, Edelstein PH, Fields BS, Geary DF, Harrison TG, et al., editors. *Legionella: state of the art 30 years after its recognition*. Washington, DC: ASM Press; 2006. p. 367−76.
25. Dykhuizen DE, Brisson D, Sandigursky S, Wormser GP, Nowakowski J, Nadelman RB, et al. The propensity of different *Borrelia burgdorferi* sensu stricto genotypes to cause disseminated infections in humans. *Am J Trop Med Hyg* 2008; **78**(5):806−10.
26. Walk ST, Alm EW, Calhoun LM, Mladonicky JM, Whittam TS. Genetic diversity and population structure of *Escherichia coli* isolated from freshwater beaches. *Environ Microbiol* 2007;**9**(9):2274−88.
27. Walk ST, Alm EW, Gordon DM, Ram JL, Toranzos GA, Tiedje JM, et al. Cryptic lineages of the genus *Escherichia*. *Appl Environ Microbiol* 2009;**75**(20):6534−44.
28. Hunt DE, David LA, Gevers D, Preheim SP, Alm EJ, Polz MF. Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* 2008;**320**(5879): 1081−5.
29. Connor N, Sikorski J, Rooney AP, Kopac S, Koeppel AF, Burger A, et al. The ecology of speciation in *Bacillus*. *Appl Environ Microbiol* 2010;**76**:1349−58.
30. Doolittle WF, Zhaxybayeva O. On the origin of prokaryotic species. *Genome Res* 2009; **19**(5):744−56.

Downloaded From Vet-eBooks.com

31. Welch RA, Burland V, Plunkett 3rd G, Redford P, Roesch P, Rasko D, et al. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci USA* 2002;**99**(26):17020–4.
32. Whittam TS, Bumbaugh AC. Inferences from whole-genome sequences of bacterial pathogens. *Curr Opin Genet Dev* 2002;**12**(6):719–25.
33. Rasko DA, Rosovitz MJ, Myers GS, Mongodin EF, Fricke WF, Gajer P, et al. The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J Bacteriol* 2008;**190**(20):6881–93.
34. Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, et al. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* 2009;**5**(1):e1000344.
35. Shapiro BJ, Friedman J, Cordero OX, Preheim SP, Timberlake SC, Szabo G, et al. Population genomics of early events in the ecological differentiation of bacteria. *Science* 2012; **336**(6077):48–51.
36. Olsen MT, Nowack S, Wood JM, Becraft ED, LaButti K, Lipzen A, et al. The molecular dimension of microbial species: 3. Comparative genomics of *Synechococcus* isolates with different light responses and in situ diel transcription patterns of associated putative ecotypes in the mushroom spring microbial mat. *Front Microbiol* 2015;**6**:604.
37. Lefebure T, Stanhope MJ. Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. *Genome Biol* 2007;**8**(5):R71.
38. Paul S, Dutta A, Bag SK, Das S, Dutta C. Distinct, ecotype-specific genome and proteome signatures in the marine cyanobacteria *Prochlorococcus*. *BMC Genomics* 2010; **11**:103.
39. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc Natl Acad Sci USA* 2005;**102**(39):13950–5.
40. Cohan FM, Koeppel AF. The origins of ecological diversity in prokaryotes. *Curr Biol* 2008; **18**:R1024–34.
41. Gogarten JP, Doolittle WF, Lawrence JG. Prokaryotic evolution in light of gene transfer. *Mol Biol Evol* 2002;**19**(12):2226–38.
42. Palenik B, Ren Q, Tai V, Paulsen IT. Coastal *Synechococcus* metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity. *Environ Microbiol* 2009;**11**(2):349–59.
43. Shapiro BJ, Polz MF. Ordering microbial diversity into ecologically and genetically cohesive units. *Trends Microbiol* 2014;**22**(5):235–47.
44. Gupta A, Sharma VK. Using the taxon-specific genes for the taxonomic classification of bacterial genomes. *BMC Genomics* 2015;**16**:396.
45. Ward DM, Bateson MM, Ferris MJ, Kühl M, Wieland A, Koeppel A, et al. Cyanobacterial ecotypes in the microbial mat community of Mushroom Spring (Yellowstone National Park, Wyoming) as species-like units linking microbial community composition, structure and function. *Philos Trans Roy Soc Ser B* 2006;**361**:1997–2008.
46. Cohan FM. What are bacterial species? *Annu Rev Microbiol* 2002;**56**:457–87.
47. Staley JT. The bacterial species dilemma and the genomic-phylogenetic species concept. *Philos Trans R Soc Lond B Biol Sci* 2006;**361**(1475):1899–909.
48. Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today* 2006;**33**:152–5.
49. Gordon DM, Lee J. The genetic structure of enteric bacteria from australian mammals. *Microbiology* 1999;**145**(Pt 10):2673–82.

Downloaded From Vet-eBooks.com

50. Vernikos GS, Thomson NR, Parkhill J. Genetic flux over time in the salmonella lineage. *Genome Biol* 2007;**8**(6):R100.
51. Marri PR, Hao W, Golding GB. Gene gain and gene loss in *Streptococcus*: is it driven by habitat? *Mol Biol Evol* 2006;**23**(12):2379−91.
52. Aboal M, Werner O, García-Fernández ME, Palazón JA, Cristóbal JC, Williams W. Should ecomorphs be conserved? The case of Nostoc flagelliforme, an endangered extremophile cyanobacteria. *J Nat Conserv* 2016;**30**:52−64.
53. Dugat T, Lagrée A-C, Maillard R, Boulouis H-J, Haddad N. Opening the black box of *Anaplasma phagocytophilum* diversity: current situation and future perspectives. *Front Cell Infect Microbiol* 2015;**5**.
54. Kettler GC, Martiny AC, Huang K, Zucker J, Coleman ML, Rodrigue S, et al. Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet* 2007; **3**(12):e231.
55. Yoon CK. *Naming nature: the clash between instinct and science*. New York: Norton; 2009.
56. Sneath P, Sokal R. *Numerical taxonomy: the principles and practice of numerical classification*. San Francisco: W.H. Freeman; 1973.
57. Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, Krichevsky MI, et al. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol* 1987;**37**:463−4.
58. Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* 2005;**102**(7):2567−72.
59. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 2007;**57**(Pt 1):81−91.
60. McVean G, Awadalla P, Fearnhead P. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* 2002;**160**:1231−41.
61. Roberts MS, Cohan FM. Recombination and migration rates in natural populations of *Bacillus subtilis* and *Bacillus mojavensis*. *Evolution* 1995;**49**:1081−94.
62. Jensen PR. Linking species concepts to natural product discovery in the post-genomic era. *J Ind Microbiol Biotechnol* 2010;**37**(3):219−24.
63. Bessen DE. Population biology of the human restricted pathogen, *Streptococcus pyogenes*. *Infect Genet Evol* 2009;**9**(4):581−93.
64. Palys T, Nakamura LK, Cohan FM. Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. *Int J Syst Bacteriol* 1997;**47**(4): 1145−56.
65. Corander J, Marttinen P, Siren J, Tang J. Enhanced bayesian modelling in baps software for learning genetic structures of populations. *BMC Bioinforma* 2008;**9**:539.
66. Barraclough TG, Hughes M, Ashford-Hodges N, Fujisawa T. Inferring evolutionarily significant units of bacterial diversity from broad environmental surveys of single-locus data. *Biol Lett* 2009;**5**(3):425−8.
67. Francisco JC, Cohan FM, Krizanc D. Accuracy and efficiency of algorithms for the demarcation of bacterial ecotypes from DNA sequence data. *Int J Bioinforma Res Appl* 2014;**10**:409−25.
68. Becraft ED, Wood JM, Rusch DB, Kühl M, Jensen SI, Bryant DA, et al. The molecular dimension of microbial species: 1. Ecological distinctions among, and homogeneity within, putative ecotypes of *Synechococcus* inhabiting the cyanobacterial mat of Mushroom Spring, Yellowstone National Park. *Front Microbiol* 2015;**6**:590.

Downloaded From Vet-eBooks.com

69. Brisson D, Dykhuizen DE. ospC diversity in *Borrelia burgdorferi*: different hosts are different niches. *Genetics* 2004;**168**(2):713−22.
70. Allewalt JP, Bateson MM, Revsbech NP, Slack K, Ward DM. Effect of temperature and light on growth of and photosynthesis by *Synechococcus* isolates typical of those predominating in the octopus spring microbial mat community of Yellowstone National Park. *Appl Environ Microbiol* 2006;**72**(1):544−50.
71. Bhaya D, Grossman AR, Steunou AS, Khuri N, Cohan FM, Hamamura N, et al. Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. *ISME J* 2007;**1**(8):703−13.
72. Vos M. A species concept for bacteria based on adaptive divergence. *Trends Microbiol* 2011;**19**(1):1−7.
73. Cohan FM. Bacterial speciation: genetic sweeps in bacterial species. *Curr Biol* 2016;**26**: R112−5.
74. Bendall ML, Stevens SL, Chan LK, Malfatti S, Schwientek P, Tremblay J, et al. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. *ISME J* 2016;**10**:1589−601.
75. Kopac SM, Cohan FM. Comment on "population genomics of early events in the ecological differentiation of bacteria". *Science* 2012:336 [Internet].
76. Majewski J, Cohan FM. Adapt globally, act locally: the effect of selective sweeps on bacterial sequence diversity. *Genetics* 1999;**152**(4):1459−74.
77. Takeuchi N, Cordero OX, Koonin EV, Kaneko K. Gene-specific selective sweeps in bacteria and archaea caused by negative frequency-dependent selection. *BMC Biol* 2015;**13**:20.
78. Smith NH, Kremer K, Inwald J, Dale J, Driscoll JR, Gordon SV, et al. Ecotypes of the *Mycobacterium tuberculosis* complex. *J Theor Biol* 2006;**239**(2):220−5.
79. Mechai S, Margos G, Feil EJ, Barairo N, Lindsay LR, Michel P, et al. Evidence for host-genotype associations of *Borrelia burgdorferi* sensu stricto. *PLoS One* 2016;**11**(2): e0149345.
80. Vuong HB, Canham CD, Fonseca DM, Brisson D, Morin PJ, Smouse PE, et al. Occurrence and transmission efficiencies of *Borrelia burgdorferi* ospC types in avian and mammalian wildlife. *Infect Genet Evol* 2014;**27**:594−600.
81. Fisher ML, Castillo C, Mecsas J. Intranasal inoculation of mice with *Yersinia pseudotuberculosis* causes a lethal lung infection that is dependent on *Yersinia* outer proteins and PhoP. *Infect Immun* 2007;**75**(1):429−42.
82. Sun Y-C, Jarrett Clayton O, Bosio Christopher F, Hinnebusch BJ. Retracing the evolutionary path that led to flea-borne transmission of *Yersinia pestis*. *Cell Host Microbe* 2014; **15**(5):578−86.
83. Jorth P, Staudinger BJ, Wu X, Hisert KB, Hayden H, Garudathri J, et al. Regional isolation drives bacterial diversification within cystic fibrosis lungs. *Cell Host Microbe* 2015;**18**(3): 307−19.
84. Saiman L, Siegel J. Infection control in cystic fibrosis. *Clin Microbiol Rev* 2004;**17**(1): 57−71.
85. Warner DF, Koch A, Mizrahi V. Diversity and disease pathogenesis in *Mycobacterium tuberculosis*. *Trends Microbiol* 2015;**23**(1):14−21.
86. Gonzales-Siles L, Sjoling A. The different ecological niches of enterotoxigenic *Escherichia coli*. *Environ Microbiol* 2016;**18**(3):741−51.
87. Linke K, Ruckerl I, Brugger K, Karpiskova R, Walland J, Muri-Klinger S, et al. Reservoirs of *Listeria* species in three environmental ecosystems. *Appl Environ Microbiol* 2014; **80**(18):5583−92.

Downloaded From Vet-eBooks.com

88. Achtman M, Wagner M. Microbial diversity and the genetic nature of microbial species. *Nat Rev Microbiol* 2008;**6**(6):431−40.
89. Remonsellez F, Galleguillos F, Moreno-Paz M, Parro V, Acosta M, Demergasso C. Dynamic of active microorganisms inhabiting a bioleaching industrial heap of low-grade copper sulfide ore monitored by real-time PCR and oligonucleotide prokaryotic acidophile microarray. *Microb Biotechnol* 2009;**2**:613−24.
90. Cohan FM. Periodic selection and ecological diversity in bacteria. In: Nurminsky D, editor. *Selective sweep*. Georgetown, Texas: Landes Bioscience; 2005. p. 78−93.
91. Kumar N, Lad G, Giuntini E, Kaye ME, Udomwong P, Shamsani N,J, et al. Bacterial genospecies that are not ecologically coherent: population genomics of *Rhizobium leguminosarum*. *Open Biol* 2015;**5**:140133.
92. Liu Y, Lai Q, Goker M, Meier-Kolthoff JP, Wang M, Sun Y, et al. Genomic insights into the taxonomic status of the *Bacillus cereus* group. *Sci Rep* 2015;**5**:14082.
93. Hanage WP, Fraser C, Spratt BG. Sequences, sequence clusters and bacterial species. *Philos Trans Roy Soc Ser B* 2006;**361**(1475):1917−27.
94. Doroghazi JR, Buckley DH. Intraspecies comparison of *Streptomyces pratensis* genomes reveals high levels of recombination and gene conservation between strains of disparate geographic origin. *BMC Genomics* 2014;**15**(1):1−14.
95. Melendrez MC, Becraft ED, Wood JM, Olsen MT, Bryant DA, Heidelberg JF, et al. Recombination does not hinder formation or detection of ecological species of *Synechococcus* inhabiting a hot spring cyanobacterial mat. *Front Microbiol* 2016;**6**:1540.
96. Papke PT, Ward DM. The importance of physical isolation in microbial evolution. *FEMS Microbiol Ecol* 2004;**48**:293−303.
97. Whitaker RJ, Grogan DW, Taylor JW. Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* 2003;**301**(5635):976−8.
98. Gordon DM, Cowling A. The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. *Microbiology* 2003;**149**:3575−86.

Downloaded From Vet-eBooks.com

This page intentionally left blank

Downloaded From Vet-eBooks.com

Population Structure of Pathogenic Bacteria

3

C.P. Andam[1], L. Challagundla[2], T. Azarian[1], W.P. Hanage[1], D.A. Robinson[2]
[1]Harvard T. H. Chan School of Public Health, Boston, MA, United States; [2]University of Mississippi Medical Center, Jackson, MS, United States

1. Introduction

There is no universally accepted definition for a bacterial population. This deficiency might be viewed as a consequence of the longstanding, and sometimes contentious, discussion on the nature of species in the bacterial world.[1–6] At present, existing bacterial species labels are more often a reflection of practical necessity and arbitrary thresholds, many of which are rooted in historical methods of phenotypic characterization and without reference to any ecological or evolutionary theory. This is particularly noticeable for many bacterial pathogens that were named after the disease they cause (e.g., *Burkholderia pseudomallei* and *Burkholderia mallei*, *Neisseria meningitidis*, and *Neisseria gonorrhoeae*).[7] Resolution of the species "dilemma" is therefore essential as it has practical application to human and veterinary health, agriculture, and biosafety, but the species concept also embraces more fundamental biological questions that remain unanswered today: What are the characteristics shared between all members of a species? What is the unit of selection in bacterial evolution? How does the standard taxonomic demarcation of species reflect ecology and niche boundaries?

A frequently used method to classify bacterial isolates is multilocus sequence typing (MLST), which involves the sequencing of single-copy housekeeping gene fragments (usually seven) and using the allelic mismatches between isolates to define strains or clones.[8] This method has proven to be a popular and effective tool in microbiology, particularly in identifying clinically important lineages of pathogenic bacteria.[9] Its utility in pathogen research is largely due to its use of multiple loci, which provides greater taxonomic resolution while allowing detection of the distorting effects of recombination between strains. It is important to note that MLST genes are also not immune to recombination, which can lead to taxonomic ambiguities within and even between species. The concept of "fuzzy species" has therefore been used to describe strains containing sequences typical of more than one species and thus do not form clear and distinct sequence clusters.[10,11]

There have been great expectations that genomic sequence similarity (e.g., Average Nucleotide Identity[12]), core protein-coding genes or ribosomal MLST,[13,14] or genome BLAST distance phylogeny[15] will lead to clearly defined bacterial species and population boundaries. Genomics-based approaches to investigate population structure allow unprecedented sensitivity to detect rare genotypes and a greater resolution of

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00003-2
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

biological relationships. Genomics also has potential to further improve our understanding of the ecology of bacteria and the evolutionary processes that gave rise to the range of variable traits possessed by isolates of a species. This is particularly relevant in pathogenic bacteria as we continue to explore the underlying causes of different clinical presentations, varying severity of disease, differences between asymptomatically carried and invasive strains, and the emergence of antimicrobial-resistant and vaccine-escape strains in the population.

To attempt to infer the processes that gave rise to the immense diversity observed among bacteria may seem a daunting task. Yet, we know that the population structure of pathogenic bacteria is a balance between many different processes, including those that produce genetic variation (i.e., polymorphisms) and those that modulate the frequency of polymorphisms in a population. In this chapter, we illustrate the different processes that shape the population structure of bacterial pathogens, with emphasis on findings from large-scale genomic studies. Our discussion attempts to demonstrate how an understanding of the population structure of bacterial pathogens can be translated into practical applications.

2. Recombination in Bacterial Populations

2.1 Emergence and Persistence of Sequence Clusters

A fundamental question in bacterial speciation is how sequence clusters of closely related strains emerge and persist. These clusters arise despite significant variation in colonization, virulence, transmissibility, and other clinically important phenotypes. For relatively clonal species (i.e., low recombination), sequence clusters are characterized by sharp and unambiguous boundaries. Some of the most epidemiologically important pathogens such as *Mycobacterium tuberculosis* and *Yersinia pestis* evolve clonally, which means there is little or no recombination occurring between strains.[16]

In highly recombining species, the mechanisms of how sequence clusters pull away from the force imposed by gene flow between clusters and subsequently remain stable is not so straightforward. If the recombination rate, measured as the rate at which polymorphisms accumulate through recombination relative to mutation (r/m), varies over time or under spatially and temporally fluctuating selective pressures, this might allow some lineages to form genetically cohesive clusters.[17] Hybrid or mosaic genotypes, that is, those that are intermediate forms as a result of recombination between two species, are likely to create clusters that are not neatly partitioned into consistent identities. This has been widely observed in *Neisseria* spp. whereby some isolates cannot be unambiguously classified into either *Neisseria meningitidis* or *Neisseria lactamica* because they contain DNA from both species, reflecting a history of recombination.[10,11] On the other hand, new lineages within species can originate by recombination, as was first reported in methicillin-resistant *Staphylococcus aureus* (MRSA) where the pandemic ST239 lineage formed from a large chromosomal recombination between distinct ST8-like and ST30-like parent lineages.[18] Hence, for species and

Downloaded From Vet-eBooks.com

lineages to remain distinct, there must be barriers to genetic exchange that exist between them.

For many bacteria, the spatial isolation of lineages or allopatry can give rise to local variants. In ST239-MRSA, global patterns of diversification and dispersal reveal a strong geographical clustering at continental, national, and city scales. At the global scale, genomic comparisons indicate that three monophyletic clades, which mostly represent European, Asian, and South American populations, reflect multiple independent exports from Europe over only a few decades.[19,20] On the other hand, country-wide (e.g., within Turkey) geographical patterns are consistent with patterns of human movement between cities.[21] Sequence clusters may also form at a finer spatial scale. In the nosocomial multidrug-resistant *Enterococcus faecium*, populations from different hospitals have emerged and diversified independent of each other, and adaptation to a specific hospital niche has resulted in genetic isolation and limited gene flow between them.[22] However, a similar degree of genetic isolation in hospital-adapted populations has not been observed within *Staphylococcus epidermidis*.[23] Patient sharing between hospitals is likely to erode the separation of populations, as was observed for MRSA whereby the genetic similarity of isolates from different hospitals correlated with patient sharing between hospitals.[24,25]

Many infectious diseases, including those caused by bacterial pathogens, are constrained in their geographical distributions by ecological barriers to the spread or establishment of populations, even in the face of the homogenizing forces of human migration.[26] In some pathogens, such ecological barriers may be due to differences in host association and vector dispersal. The phylogeographical structure of bacterial pathogens is expected to mirror the structure observed in their reservoir hosts, but the association may not always be as straightforward as expected. An excellent example is the spirochete *Borrelia burgdorferi* sensu lato (*B. burgdorferi* s.l.) species complex, the causal agent of Lyme disease, which consists of >20 species. In multiple *Borrelia* species, the distribution and migration of both host and vector appear to greatly determine the geographical distribution of the bacterium. In a 2011 study, limited geographical structuring between countries in populations of *Borrelia* spp. associated with birds (*Borrelia garinii* and *Borrelia valaisiana*) was observed, likely as a result of higher rates of migration of the host, but there exists a strong signal in geographical structure in the bacterium associated with small mammals (*Borrelis afzelii*).[27] In another study, *B. burgdorferi* populations from Europe and the United States isolated from human patients exhibit an overlap of sequence types (STs), which is in contrast to populations found in tick vectors.[28]

In some cases, strains coexisting within the same host species remain isolated despite having ample opportunity for genetic exchange. This has been observed in *Campylobacter jejuni*, a gut colonizer of many animal species and a causal agent of gastroenteritis in humans.[29] Despite having a high degree of niche overlap and the ability to readily recombine with each other in vitro, two generalist lineages possess separate gene pools. A cryptic ecological barrier within the host appears to exist between the two lineages, which likely explains the lack of gene flow between them. The exact nature of this barrier, however, remains unclear. In other cases, genetic barriers can arise that prevent recombination between strain lineages; this has been reported for

Downloaded From Vet-eBooks.com

B. pseudomallei where lineage-specific restriction-modification systems, carried on mobile genetic elements, allow recombination within but not between clades.[30]

Bacterial pathogens form and maintain distinct species and sequence clusters when barriers to gene flow exist. In many instances, these barriers are leaky or the mechanisms underlying cluster formation may not be apparent. Sequence clusters can be observed at different scales, from global geographical distributions, to within host and vector species. These sequence clusters may fuse through recombination, or remain distinct from each other because of ecological or genetic barriers to recombination. The question as to whether these clusters are biologically meaningful must be addressed on a case-by-case basis.

2.2 Heterogeneity in Recombination

Population genomic studies of hundreds and even thousands of bacterial pathogens reveal a remarkable level of variation in recombination rates and patterns across very closely related lineages. Hence, this variation adds another layer of complexity that may conflict with blanket definitions of bacterial species and populations. Moreover, this variation may be associated with certain clinically relevant phenotypes, such as antibiotic resistance. In the pneumococcus, for example, hyper-recombinant populations have been reported to exhibit higher levels of resistance to multiple antibiotics.[31] In *Acinetobacter baumannii*, transformation experiments indicate that multidrug resistance (MDR) evolved faster in recombining, functionally diverse populations within only a few generations compared to nonrecombining populations.[32] The authors also show that the average fitness of MDR genotypes and their spread depended on whether they arose by mutation or recombination.

Recombination rates can vary significantly within a species. The genetically promiscuous pneumococcus exhibits a dramatic range in recombination rates. Early studies using MLST detected differences in r/m between capsular serotypes of the same serogroup.[33] Surveys using genome sequences from numerous serotypes of pneumococci have estimated a range of r/m values, from 0.06 to 34.06.[34] Other bacterial pathogens also exhibit such variation. The causative agent of listeriosis, a serious infection caused by eating food contaminated with *Listeria monocytogenes*, consists of at least four evolutionary lineages. Higher recombination rates are found in lineage II strains, which are widespread in natural and farm environments, and are also commonly isolated from animal listeriosis cases. The higher recombination rates in lineage II may contribute to its adaptation to diverse environments and hosts. In contrast, lineage I, the predominant cause of human listeriosis outbreaks, is largely clonal.[35,36]

Different strains also vary in terms of how often they donate or receive recombined DNA, and this may greatly influence the population structure and dynamics of pathogens. These biases, wherein some lineages act as frequent donors while others prefer to receive DNA more often, can create highways of gene exchange. In the pneumococcus, certain sporadically occurring, nonencapsulated lineages (i.e., ST1106) appear to recombine more often than others, potentially forming a hub for gene flow that is an important source of genetic diversity for the wider population.[37] This characteristic is

Downloaded From Vet-eBooks.com

not simply due to the absence of a polysaccharide capsule, which may act as a physical barrier to the entry of exogenous DNA, because other nonencapsulated lineages (i.e., ST344 and ST448) show no significant difference in r/m compared to encapsulated lineages.[34,38] In the emerging opportunistic pathogen *Mycobacterium abscessus*, which causes lung diseases in immune-compromised individuals, asymmetrical gene flow also occurs among three subspecies; subspecies *Mycobacterium bolletii* appears to donate more often to *Mycobacterium massiliense* than the other way around.[39] In the human skin commensal and opportunistic pathogen *S. epidermidis*, one population (i.e., genetic cluster 3) appears to receive DNA from all other clusters but does not donate DNA to these other clusters.[23]

Recombination also does not occur at constant rates across the genome, as "recombination hot spots" are consistently found. In the pneumococcus, the common hot spots contain genes encoding cell surface antigens, such as pneumococcal surface protein A (*pspA*) and pneumococcal surface protein C (*pspC*), as well as antibiotic resistance, such as penicillin-binding proteins (*pbp2x*, *pbp1a*, *pbp2b*) and dihydrofolate reductase (*folA*).[37] These hot spots are likely driven by selective pressure due to host immunity and clinical intervention.[37] In *C. jejuni* genomes, three recombination hot spots have been identified.[40] More than half of the genes in these hot spots are related to membrane proteins that are crucial for host interaction, colonization, and adhesion to intestinal epithelial cells, and are likely under diversifying selection as a response to the host immune system.

In *S. aureus*, recombination hot spots appearing over a megabase scale have been linked to large chromosomal replacements that have occurred around the origin of replication (e.g., Refs. 18,41), and hot spots appearing over a kilobase scale occur at insertion sites of mobile genetic elements.[42] The size of DNA fragments being recombined also varies dramatically in the pneumococcus, with two main types identified: (1) micro-recombination, which involves the frequent replacement of single, short DNA fragments with mean size of 27−580 bp and (2) macro-recombination, though rarer, involves the acquisition of multiple, long fragments with mean size 8800−14,000 bp and is associated with major phenotypic changes.[43] While the mechanism underlying these different import size distributions is unknown, it has been hypothesized to be driven by the saturation of the mismatch repair system.[43]

2.3 The Structure of the Pan-Genome of Species and Populations

Whole-genome sequencing (WGS) provides the means to characterize the full range of diversity within a species. These studies reinforce the concept that a single strain cannot fully represent the rest of the species. A more profound understanding of the species' total genetic makeup is essential and the concept of the pan-genome paves the way to achieve this. The full complement of genes in a given species or population under consideration is referred to as the pan-genome. It includes both the genes present in all strains (core genome) and the genes present only in some strains of a species (accessory, flexible, or dispensable genome). In the following paragraphs, we provide some examples of how the pan-genome may relate to bacterial population structure.

Downloaded From Vet-eBooks.com

In a classic study of three *Escherichia coli* genomes, each one having a different pathogenic characteristic (nonpathogenic strain K12, uropathogenic strain CFT073, enterohemorrhagic strain EDL933), only 39.2% of the combined set of genes is common to all three strains.[44] While this backbone of common genes has largely evolved by vertical inheritance, the remaining genes unique to each strain have been acquired by frequent gene gain and loss. More recent genomic studies of *E. coli* indicate a more remarkable level of intraspecific diversity. Comparison of the gene content of 61 *E. coli* strains shows a total of 15,741 gene families comprising the pan-genome, but only 933 (6%) are present in every genome in the sample.[45] As a typical *E. coli* genome contains 5000 genes, the accessory genome therefore makes up $\geq$ 90% of the pan-genome and approximately 80% of an individual genome.[45] Similar findings were reported in a study of 186 *E. coli* genomes with only 1072 gene families present in all strains and pan-genome size of 16,373 gene families.[46] Despite the relatively small size of the core genome, the same study also reported that the core genome tree shows a clear demarcation between the seven *E. coli* phylogroups, which is mostly recapitulated in the accessory gene presence−absence tree. Pan-genome analysis has also clarified the relationship of *E. coli* and its close relative *Shigella*. The genetic repertoire of the two species is very similar, indicating a common gene pool with no evidence for *Shigella*-specific genes.[47] Hence, *Shigella*, which cannot be justified as being its own genus based on its unique gene content or its paraphyletic location within *E. coli*, is another example of a historical anomaly in bacterial taxonomy.

Many species of pathogenic bacteria can be isolated from environmental sources. Pan-genome analysis is an excellent approach to uncover the genetic differences between clinical and environmental populations and to understand the evolution of pathogenicity in a species. One example is *P. aeruginosa*, an opportunistic pathogen able to thrive in diverse ecological niches. It is also one of the leading causes of nosocomial infections and an infectious agent for cystic fibrosis patients. *P. aeruginosa* has a large core genome, which in one study was estimated at 5233 gene families, or about 88% of the average genome.[48] Remarkably, the gene content is extremely well conserved between environmental (plant, animal, water) and clinical populations.[49] Hence, environmental isolates are likely to contain the necessary traits to become pathogenic to humans. This is in contrast with the population structure of *Vibrio vulnificus*, a halophilic bacterium typically found in marine environments and is also an important pathogen that causes fatal bloodstream and skin infections. Phylogenies based on the pan-, core, and accessory genomes show similar clustering patterns with two major clusters that mostly represent the clinical and environmental populations.[50] While we may consider the accessory genomes of *V. vulnificus* as a reflection of adaptation to diverse environments, the case of *P. aeruginosa* may suggest a different mechanism underlying its population structure.

Pan-genome analysis of bacterial pathogens brings a deeper recognition of the enormous genetic diversity in bacterial pathogens. At the same time, it has also demonstrated that species and population boundaries may not be resolved as quickly and as unambiguously as we might expect. While the terms pan-, core, and accessory

Downloaded From Vet-eBooks.com

genomes may be useful in population genomic studies, we have to keep in mind that their gene contents are partially a result of the sampling criteria used and may not necessarily reflect the ecology of the population under study. An important concept that has been put forward is the existence of a horizontal gene pool comprising niche-adaptive genes from the flexible genome.[51] The dynamics of this gene pool drive population structuring through an initial ecological separation, which reduces the rate of gene flow between nascent populations. Later on, population-specific mutations will accumulate between them, further reducing gene flow, and will lead to the formation of distinct, ecologically differentiated sequence clusters.[51]

2.4 Gene Flow Across Species Boundaries

The ability of some bacterial pathogens to obtain genetic material beyond the bounds of named species suggests that their accessory gene pool is extremely large. This is particularly worrisome because the constant barrage of clinical interventions imposed on bacterial populations is an opportunity for them to alter their genomes and rapidly adapt to the interventions. The pneumococcus, again, is an excellent example of this adaptability. The pneumococcus exchanges genes with closely related streptococcal species, such as *S. mitis*, *S. infantis*, and *S. oralis*.[52] Genomic comparisons between these species indicate that about 30% of biallelic polymorphisms in the pneumococcus are also polymorphic in *S. mitis*. These polymorphisms include those in pathogenicity genes, and their existence supports the notion that *S. mitis* is a genetic reservoir for the pneumococcus.[53] Alternatively, some fraction of these shared polymorphisms could represent polymorphisms from an ancestral population; this possibility has important implications for population genetic modeling but is seldom considered with bacteria. On the other hand, some bacteria appear to be more discriminating in exchanging genes with other species. Homologous recombination between *S. aureus* and *S. epidermidis* is generally limited, but transfer of mobile genetic elements between species is known to occur.[54,55]

Recombination probably influences the diversification of all bacterial pathogens, even those, such as *E. coli*, that provided early examples of the clonal nature of bacteria.[56] However, recombination's contribution to diversity, r/m, is highly variable not only across species but also within species. Adding to this variation is that recombination may vary over time or across the phylogeny. In *E. coli*, for example, different recombination rates emerge at different clustering levels, such as phylogenetic group, lineage, and clonal complex.[57] Recombination rates can also differ across a bacterial chromosome and the size of the recombination events may reflect different mechanisms. Recombination has a complex relationship to population structure and ecology. Frequent transfers between sequence clusters (be they species or populations) can be a homogenizing force, whereas rare transfers between clusters can be a diversifying force. The balance between these two forces, in addition to the generation of polymorphisms due to mutation, will greatly influence the population structure of bacterial pathogens.

Downloaded From Vet-eBooks.com

3. Evolutionary Processes Shape Intra- and Interhost Bacterial Population Structure

3.1 Intrahost Evolution: A Snapshot of Larger-Scale Population Dynamics

As with epidemiological study samples, bacterial populations may be defined by the scale of observation. Prior to the availability of low-cost WGS, the scale of observation was most often interhost and used molecular typing methods with substantially poorer resolution. However, the fine-grained typing resolution provided by WGS has allowed investigators to narrow the scale of observation to the level of an individual host. As a result, we have gained a better understanding of population dynamics of bacterial pathogens within and between hosts.

For some time, it has been known that the assumption of a genetically and phenotypically homogenous bacterial population within an individual host was an oversimplification. Throat carriage of multiple pulsed-field gel electrophoresis types of *Haemophilus influenzae* (43%) and *S. pneumoniae* (5%) has been observed,[58] consistent with older studies that showed the possibility of carriage of multiple serotypes of these species. Nasal carriage of multiple *spa* types has also been observed in *S. aureus* (11%).[59] These examples indicate the presence of intrahost diversity. The limitations of low resolution typing methods and the practice of isolating a single strain from a specimen can potentially be overcome by WGS, thus allowing the full breadth of intrahost diversity to be revealed. It has since been shown that intrahost diversity can vary tremendously as a result of multiple factors including the bottleneck size of the bacterial population acquired during a transmission event, mutation rate, colonization and/or infection duration, and selective pressures, such as host-immunity or treatment with antibiotics.[60] Furthermore, simulation studies have demonstrated that ignoring intrahost diversity during epidemic investigations hinders accurate reconstruction of interhost transmission networks.[61] Studies during 2013−16 have leveraged the increased affordability of WGS to assess these dynamics by sequencing multiple bacterial isolates during colonization and/or infection both using cross-sectional and longitudinal sampling.[62,63,64] This has revealed that intrahost bacterial populations are indeed measurably evolving, a distinction that had previously been reserved for fast-evolving viruses, such as hepatitis or HIV.[65]

Cross-sectional snapshots of intrahost bacterial populations have shown a range of nucleotide diversity. Studies of *S. aureus* colonization, for example, have found cohesive populations of bacteria separated by up to about 40 single nucleotide polymorphisms (SNPs) with no apparent subpopulation structure among body sites sampled.[62,63,66] In longitudinal studies of *S. aureus* carriage, significant intrahost fluctuations in effective population size (N_e) have been inferred[62] as well as an accumulation of mutations preceding the transition from colonization to disease.[66] It was also shown that periods of high mutation are punctuated by periods when no mutations were observed, suggesting intrahost evolutionary rate heterogeneity.[66]

Downloaded From Vet-eBooks.com

Estimates of intrahost mutation rates are also variable among species and have ranged from as high as about 30 SNPs/year for *H. pylori*[67] to as low as 0.5 SNPs/year for *M. tuberculosis*,[68] with other species including *Klebsiella pneumoniae* (10 SNPs/year)[69] and *S. aureus* (8 SNPs/year)[66] falling in between this range. These mutations may have various fitness effects. On the comparatively shorter timescale of intrahost evolution, the effects of random genetic drift and incomplete purifying selection often result in an excess of nonsynonymous mutations. Over time, purifying selection would likely remove weakly deleterious mutations.

Estimates of intrahost diversity must be made in light of several considerations. Primarily, the duration of colonization and transmission bottleneck size is often unknown in these studies. The size of the bacterial population acquired during a transmission event (i.e., bottleneck size) would greatly impact the diversity of the founding intrahost population, and assuming a constant evolutionary rate, diversity would increase over time. Secondly, bioinformatics pipelines for identifying SNPs in bacterial populations are not standardized and can vary significantly between studies. Lastly, multiple transmission events of a closely related strain are likely within highly connected transmission chains or high transmission settings (e.g., within a household), potentially resulting in inflated intrahost diversity estimates. Taken together in consideration of strain dynamics (e.g., hypermutators and highly recombinant strains), a single snapshot of the intrahost population may not accurately reflect the demographic history of the pathogen, and it is for this reason why a single cut-off value of SNP difference should not be used to infer interhost transmission events. Sampling multiple bacterial isolates over time better captures the intrahost population dynamics.

Bacterial populations are often thought of on a larger scale than an individual host. However, as we have illustrated, technological advances in WGS combined with epidemiological studies sampling multiple bacterial isolates from individuals over time have redefined our concept of a bacterial population. Within a single patient that is harmlessly colonized by bacteria or that is being treated for a bacterial infection, we can observe extensive polymorphisms among the bacterial isolates and can infer these to be the result of the gamut of evolutionary processes. In essence, intrahost studies serve as a model for gaining a better understanding of population structure on a larger scale.

3.2 Interhost Evolution and Population Structure

Intrahost populations are acquired through person-to-person, zoonotic, or environmental transmission events, which can collectively be termed "interhost." The acquired bacterial population arises from a subpopulation within a species, which may possess temporal, geographical, environmental, or host-specific structure. The evolutionary forces shaping the population structure are indeed the same as those acting on intrahost populations, and the way in which they do so has long been investigated by researchers in microbiology, population genetics, and epidemiology. Now, with the widespread application of WGS and an assortment of phylogenetic and population

Downloaded From Vet-eBooks.com

genetic analysis tools, the demographic history and transmission dynamics of bacteria can be studied at an unprecedented level of resolution. For example, studies reported during 2010−14 have used these methods to investigate clonal emergence and intercontinental spread,[19,70] transmission within the community or in high-risk contact networks,[71,72] and to track geographical origin.[73]

Studies of *Vibrio cholerae* provide an excellent example of what can be learned about interhost evolution and population structure. *V. cholerae* serogroup O1, consisting of "classical" and "El Tor" biotypes, causes epidemic cholera and is the most epidemiologically important lineage. Genomic studies have shown that serogroup O1 is clonal and that both the O1 antigen as well as the genomic region coding cholera toxin were acquired through horizontal gene transfer. Non-O1 strains, commonly found in the environment, are considerably more diverse and do not demonstrate a cohesive population structure, although this may be a result of incomplete sampling.[74]

Molecular clock analysis (discussed further in the following section) using time-dated genomes found that the current seventh pandemic of O1 cholera originated in 1950s in the Bay of Bengal and caused at least three overlapping epidemic waves.[75] Phylogeographical analysis further illustrated that the first of these waves spread globally, while the second and third waves were geographically isolated. Further assessment showed that recombination played a significant role in the diversification of each wave, which were preceded by considerable population bottlenecks. Specifically, the acquisition of SXT/R391 integrative and conjugative elements through recombination conferred tetracycline and furazolidone resistance and may have been adaptive, as the clinical use of these antibiotics for cholera preceded the development of resistance by roughly 15 years.[75] This analysis also demonstrated the importance of longitudinal sampling, which for cholera is more complete than some other pathogens. Prior to 2014, the oldest cholera isolates dated to the 1930s; however, as reported in 2014, a study of a second pandemic strain obtained from the preserved intestine of a patient who died in the Philadelphia cholera epidemic of 1849 extended the study of the demographic history of cholera an additional century.[76] The authors of this study reported that the 1849 strain differed by only 203 SNPs from the classical O395 strain isolated in 1965, but was missing three genomic islands, leading to a recalibration of the evolutionary rate for *V. cholerae*.[76]

In October 2010, a cholera epidemic struck the already earthquake-torn country of Haiti, which had been cholera-free for over 100 years.[77] While this epidemic caused significant morbidity and mortality, it also provided a unique opportunity to study the emergence of *V. cholerae* in a new setting. Phylogenetic analysis was used to trace the introduction of cholera to a Nepalese garrison on a United Nations peacekeeping duty in Haiti.[73,78] Isolates from ongoing outbreaks in Bangladesh and Nepal formed a monophyletic clade with those from Haiti, with several Nepalese strains differing from the Haitian epidemic clone by only 1−2 SNPs. Combined with epidemiological evidence, the Nepalese origin of the Haitian epidemic appears the most parsimonious. At the time of introduction, the phylogeny of the Nepalese cholera epidemic was divided into four distinct clades.[78] The introduction to Haiti resulted in a substantial population bottleneck followed by rapid population expansion and has been

Downloaded From Vet-eBooks.com

documented in other studies.[79,80] Even within a relatively short time span, a progressive accumulation of mutation was observed with evidence of diversifying selection.[80] The epidemic quickly spread to all regions of the county as well as to neighboring Dominican Republic, and environmental surveillance subsequently isolated the epidemic clone from the several Haitian rivers and estuaries.[81] Additionally, a strain was identified that had acquired a multidrug-resistant IncA/C plasmid putatively from a member of the family Enterobacteriaceae.[82] In all, these examples demonstrate how WGS and phylogenetic analysis can facilitate the understanding of historical patterns of spread, the role of advantageous recombinations and mutations, and the temporal−spatial origin and spread of a pathogen.

4. Genomic Analysis Tools for Studying Bacterial Population Structure

One approach for inferring bacterial population structure within frequently recombining bacterial species is to assign strains to sequence clusters based on allele frequencies. Two widely used methods for this "population assignment" approach with bacteria include BAPS[83] and STRUCTURE.[84] Both methods infer the number of populations present in a sample. In addition, both methods allow a proportion of an individual's genome to derive from different ancestral populations, and both methods can account for linkages between SNPs. However, BAPS and STRUCTURE implement different Bayesian models with different computational efficiency. Nonmodel-based approaches, which tend to be less computationally intensive than the model-based approaches, are also used to infer population structure. These include standard multivariate analysis methods, such as Principal Component Analysis (PCA) and the combination of discriminant analysis with PCA.[85] fineSTRUCTURE, which uses information on haplotypes, is a newer method that attempts to unify the model- and nonmodel-based approaches for inferring population structure.[86] In 2013, Yahara et al.[87] applied this method with genome sequences to identify population structure within *H. pylori*. They distinguished more populations than by using either phylogenetic methods or STRUCTURE based on MLST. In addition, this method was applied with genome sequences to delineate populations of *K. pneumoniae*.[88] Three major lineages, corresponding to previously described phylogroups within *K. pneumoniae*, as well as numerous minor lineages within the principal human pathogenic lineage, were reported. In addition, the study provided genomic support for the notion that these three major lineages represent separate species.[88]

An abundance of methods are available for inferring phylogenies, which provide understanding of the tree-like relationships between strains. Phylogenetic reconstructions can also be used as part of the inference of rates and paths of transmission[89,90,91] and to infer most recent common ancestors.[79] Maximum likelihood (ML) methods, which assume independence between all sites in the alignment, are commonly used to infer phylogenies from samples of bacterial genomes. ML methods first select a topology and then calculate the likelihood of the data given the fit to the proposed

Downloaded From Vet-eBooks.com

topology and nucleotide substitution model; the process is then iterated and the topology with the highest likelihood is selected. PhyML[92] and RAxML[93] are commonly used ML methods. ExaML is an implementation of RAxML reported in 2015 with improved efficiency for use with computing clusters.[94] FastTree is another ML method that was developed to increase the scalability of analysis to allow ever-larger samples of genomes.[95] The tradeoff in computational speed is in the accuracy of the phylogeny; however, with datasets of thousands of bacterial genomes, one may be more interested in the relationships between larger clades, not fine-scaled resolution.

Other phylogenetic approaches have been developed that can simultaneously model nucleotide substitution, population demography, and genealogy. BEAST (Bayesian Evolutionary Analysis Sampling Trees)[96] is a popular software for this purpose. It can generate time-calibrated phylogenies and can assess the demographic history of pathogens, including changes in effective population size over time.[97] As one example of the use of this approach, McAdam et al.[98] established a time frame for the emergence and spread of MRSA clones from the CC30 lineage. Their study implicated patient referrals from hospitals in metropolitan areas to more regional settings in the United Kingdom as potentially important transmission pathways.

The importance of recombination in shaping bacterial population structure has been extensively discussed throughout this chapter. The presence of recombinant DNA fragments can hinder phylogenetic and demographic inference by not only producing phylogenetic inconsistency, in which different sites support different tree topologies, but also by distorting branch lengths.[99,100] Therefore, it is imperative to identify recombinations in bacterial genome sequences, preferably in a computationally efficient manner. This challenge has been partially addressed by several methods. BratNextGen[101] utilizes a Bayesian model to identify distinct clusters of taxa at varying distances along an alignment. A permutation resampling of the SNPs in the alignment is used to estimate the statistical significance of the identified recombinant segments.[101] ClonalframeML utilizes an inferred ML phylogeny (e.g., generated by PhyML), and estimates ancestral sequences at internal nodes on the phylogeny as well as branch lengths corrected for recombination events and recombination parameters.[102] This approach has a computational advantage over an earlier Bayesian implementation that makes it efficient for analyzing hundreds of genomes. Gubbins is a method that identifies recombinant fragments as loci with high densities of SNPs, simultaneously with the generation of an ML phylogeny from the remaining loci.[103] Finally, STARRInIGHTS combines information on both SNP distribution and tree topology, which allows for the detection of the breakpoints between recombination blocks.[104] Since these different methods may not always agree about which SNPs are recombinant, sensitivity analyses and comparisons of different methods may be a necessary component of study.

Population genetic summary statistics have been widely used to characterize patterns of genetic variation in bacteria. To calculate summary statistics from large samples of genomes, standalone software, such as VariScan,[105] and R packages, such as PopGenome,[106] are available. Comparisons of sequence polymorphism within populations and divergence between populations can lead to the discovery of genetic loci

Downloaded From Vet-eBooks.com

that may have been affected by natural selection. For example, in a study of *M. tuberculosis* reported in 2015[107] nucleotide diversity was quantified within and between hosts in order to identify atypical loci, relative to the diversity of all loci in the genome, which may contribute to host adaptation. They found low values for Tajima's D in genes associated with cell envelope lipids, leading to the hypothesis that such genes may impact intrahost adaptation.

Such "outlier detection" methods work from the premise that natural selection will nudge genetic variation at loci affected by the selection to the extremes of the empirical distribution of genetic variation in the sample. However, it is crucial to note that strain sampling procedures can also nudge the empirical distribution one way or another, as can purely neutral demographic processes. To address these issues in a study focused on identifying candidate targets of balancing selection in *S. aureus*, Thomas et al.[108] used approximate Bayesian computation (ABC) methods. The simplest form of ABC involves simulation and rejection sampling to fit a model to data.[109] Thomas et al.[108] used ABC to develop several null models, which fit their data better than did the standard neutral model, in order to judge the unusualness of genetic variation at different loci. Summary statistics were selected to be those most sensitive to balancing selection (e.g., Tajima's D and the ratio of intraspecific polymorphism to interspecific divergence). They discovered that the master virulence gene regulator, *agr*, has hallmark characteristics of balancing selection, including unusually elevated polymorphism that reflects distinct allelic groups with potentially distinct functions. In addition, strong signals for balancing selection were detected in genes that may do double-duty in providing resistance to glyco- and lipopeptide antibiotics and cationic antimicrobial peptides from the host immune system.

5. Conclusions

Since 2010, considerable progress has been made in inferring the relative contributions of the different evolutionary processes that shape the population structure of pathogenic bacteria. Analyses of genome sequences from samples of hundreds and even thousands of bacterial isolates have allowed the precise identification of sequence clusters that sometimes correspond to lineages within species and sometimes may deserve their own species labels. Moreover, these analyses have provided precise estimates of how much recombination and mutation occur within and between these clusters, and they have provided understanding of the adaptations that drive the associations of bacterial pathogens to their hosts and ecological niches. Future work would be needed to precisely map genotype−phenotype associations, which has long been used in human genetics, but is still at a nascent stage in microbiology. These association studies can be used to determine the genetic factors underlying the heterogeneity in antimicrobial resistance, invasive disease potential, and host specificity of bacterial pathogens. This information may increasingly direct how we manage pathogenic bacteria.

Downloaded From Vet-eBooks.com

References

1. Godreuil S, Cohan F, Shah H, Tibayrenc M. Which species concept for pathogenic bacteria? An E-debate. *Infect Genet Evol* 2005;**5**:375−87.
2. Doolittle WF, Papke RT. Genomics and the bacterial species problem. *Genome Biol* 2006; **7**:116.
3. Cohan FM, Perry EB. A systematics for discovering the fundamental units of bacterial diversity. *Curr Biol* 2007;**17**:R373−86.
4. Fraser C, Alm EJ, Polz MF, Spratt BG, Hanage WP. The bacterial species challenge: making sense of genetic and ecological diversity. *Science* 2009;**323**:741−6.
5. Bapteste E, Boucher Y. Epistemological impacts of horizontal gene transfer on classification in microbiology. *Methods Mol Biol* 2009;**532**:55−72.
6. Shapiro BJ, Polz MF. Microbial speciation. *Cold Spring Harb Perspect Biol* 2015;**7**: a018143.
7. Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, et al. Re-evaluating prokaryotic species. *Nat Rev Microbiol* 2005;**3**:733−9.
8. Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* 1998;**95**:3140−5.
9. Feil EJ, Smith JM, Enright MC, Spratt BG. Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data. *Genetics* 2000;**154**: 1439−50.
10. Hanage WP, Fraser C, Spratt BG. Fuzzy species among recombinogenic bacteria. *BMC Biol* 2005;**3**:6.
11. Corander J, Connor TR, O'Dwyer CA, Kroll JS, Hanage WP. Population structure in the *Neisseria* and the biological significance of fuzzy species. *J R Soc Interface* 2012;**9**: 1208−15.
12. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, et al. Microbial species delineation using whole genome sequences. *Nucl Acids Res* 2015;**43**:6761−71.
13. Bennett JS, Jolley KA, Earle SG, Corton C, Bentley SD, Parkhill J, et al. A genomic approach to bacterial taxonomy: an examination and proposed reclassification of species within the genus *Neisseria*. *Microbiology* 2012;**158**:1570−80.
14. Jolley KA, Maiden MCJ. Using multilocus sequence typing to study bacterial variation: prospects in the genomic era. *Future Microbiol* 2014;**9**:623−30.
15. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bionf* 2013; **14**:60.
16. Godreuil S, Renaud F, Choisy M, Depina JJ, Garnotel E, Morillon M, et al. Highly structured genetic diversity of the *Mycobacterium tuberculosis* population in Djibouti. *Clin Microbiol Infect* 2010;**16**:1023−6.
17. Cox MP, Holland BR, Wilkins MC, Schmid J. Reconstructing past changes in locus-specific recombination rates. *BMC Genet* 2013;**14**:11.
18. Robinson DA, Enright MC. Evolution of *Staphylococcus aureus* by large chromosomal replacements. *J Bacteriol* 2004;**186**:1060−4.
19. Harris SR, Feil EJ, Holden MTG, Quail MA, Nickerson EK, Chantratita N, et al. Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 2010;**327**: 469−74.

Downloaded From Vet-eBooks.com

20. Smyth DS, McDougal LK, Gran FW, Manoharan A, Enright MC, Song JH, et al. Population structure of a hybrid clonal group of methicillin-resistant *Staphylococcus aureus* ST239-MRSA-III. *PLoS One* 2010;**5**:e8582.
21. Castillo-ramirez S, Corander J, Marttinen P, Aldeljawi M, Hanage WP, Westh H, et al. Phylogeographic variation in recombination rates within a global clone of Methicillin-Resistant *Staphylococcus aureus* (MRSA). *Genome Biol* 2012;**13**:R126.
22. Willems RJL, Top J, van Schaik W, Leavis H, Bonten M, Sirén J, et al. Restricted gene flow among hospital subpopulations of *Enterococcus faecium*. *mBio* 2012;**3**. e00151−12.
23. Thomas JC, Zhang L, Robinson DA. Differing lifestyles of *Staphylococcus epidermidis* as revealed through Bayesian clustering of multilocus sequence types. *Infect Genet Evol* 2014;**22**:257−64.
24. Ke W, Huang SS, Hudson LO, Elkins KR, Nguyen CC, Spratt BG, et al. Patient sharing and population genetic structure of methicillin-resistant *Staphylococcus aureus*. *Proc Natl Acad Sci USA* 2012;**109**:6763−8.
25. Chang H-H, Dordel J, Donker T, Worby CJ, Feil EJ, Hanage WP, et al. Identifying the effect of patient sharing on between-hospital genetic differentiation of methicillin-resistant *Staphylococcus aureus*. *Genome Med* 2016;**8**:18.
26. Murray KA, Preston N, Allen T, Zambrana-Torrelio C, Hosseini PR, Daszak P. Global biogeography of human infectious diseases. *Proc Natl Acad Sci USA* 2015;**112**: 12746−51.
27. Vollmer SA, Bormane A, Dinnis RE, Seelig F, Dobson ADM, Aanensen DM, et al. Host migration impacts on the phylogeography of Lyme Borreliosis spirochaete species in Europe. *Environ Microbiol* 2011;**13**:184−92.
28. Jungnick S, Margos G, Rieger M, Dzaferovic E, Bent SJ, Overzier E, et al. *Borrelia burgdorferi sensu stricto* and *Borrelia afzelii*: population structure and differential pathogenicity. *Int J Med Microbiol* 2015;**305**:673−81.
29. Sheppard SK, Cheng L, Méric G, De Haan CPA, Llarena AK, Marttinen P, et al. Cryptic ecology among host generalist *Campylobacter jejuni* in domestic animals. *Mol Ecol* 2014; **23**:2442−51.
30. Nandi T, Holden MTG, Didelot X, Mehershahi K, Boddey JA, Beacham I, et al. *Burkholderia pseudomallei* sequencing identifies genomic clades with distinct recombination accessory and epigenetic profiles. *Genome Res* 2015;**25**:129−41.
31. Hanage WP, Fraser C, Tang J, Connor TR, Corander J. Hyper-recombination diversity and antibiotic resistance in pneumococcus. *Science* 2009;**324**:1454−7.
32. Perron GG, Lee AEG, Wang Y, Huang WE, Barraclough TG. Bacterial recombination promotes the evolution of multi-drug-resistance in functionally diverse populations. *Proc Biol Sci* 2012;**279**:1477−84.
33. Robinson DA, Briles DE, Crain MJ, Hollingshead SK. Evolution and virulence of serogroup 6 pneumococci on a global scale. *J Bacteriol* 2002;**184**:6367−75.
34. Croucher NJ, Finkelstein JA, Pelton SI, Mitchell PK, Lee GM, Parkhill J, et al. Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nat Genet* 2013;**45**: 656−63.
35. den Bakker HC, Didelot X, Fortes ED, Nightingale KK, Wiedmann M. Lineage specific recombination rates and microevolution in *Listeria monocytogenes*. *BMC Evol Biol* 2008; **8**:277.
36. Orsi RH, Bakker HCD, Wiedmann M. *Listeria monocytogenes* lineages: genomics evolution ecology and phenotypic characteristics. *Int J Med Microbiol* 2011;**301**:79−96.

Downloaded From Vet-eBooks.com

37. Chewapreecha C, Harris SR, Croucher NJ, Turner C, Marttinen P, Cheng L, et al. Dense genomic sampling identifies highways of pneumococcal recombination. *Nat Genet* 2014; **46**:305−9.
38. Hilty M, Wüthrich D, Salter SJ, Engel H, Campbell S, Sá-Leão R, et al. Global phylogenomic analysis of nonencapsulated *Streptococcus pneumoniae* reveals a deep-branching classic lineage that is distinct from multiple sporadic lineages. *Genome Biol Evol* 2014;**6**: 3281−94.
39. Sapriel G, Konjek J, Orgeur M, Bouri L, Frézal L, Roux A-L, et al. Genome-wide mosaicism within *Mycobacterium abscessus*: evolutionary and epidemiological implications. *BMC Genomics* 2016;**17**:1−16.
40. Yahara K, Didelot X, Ansari MA, Sheppard SK, Falush D. Efficient inference of recombination hot regions in bacterial genomes. *Mol Biol Evol* 2014;**31**:1593−605.
41. Nimmo GR, Steen JA, Monecke S, Ehricht R, Slickers P, Thomas JC, et al. ST2249-MRSA-III: a second major recombinant methicillin-resistant *Staphylococcus aureus* clone causing healthcare infection in the 1970s. *Clin Microbiol Infect* 2015;**21**: 444−50.
42. Everitt RG, Didelot X, Batty EM, Miller RR, Knox K, Young BC, et al. Mobile elements drive recombination hotspots in the core genome of *Staphylococcus aureus*. *Nat Comms* 2014;**5**:3956.
43. Mostowy R, Croucher NJ, Hanage WP, Harris SR, Bentley S, Fraser C. Heterogeneity in the frequency and characteristics of homologous recombination in pneumococcal evolution. *PLoS Genet* 2015:101−15.
44. Welch RA, Burland V, Plunkett G, Redford P, Roesch P, Rasko D, et al. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc Natl Acad Sci USA* 2002;**99**:17020−4.
45. Lukjancenko O, Wassenaar TM, Ussery DW. Comparison of 61 sequenced *Escherichia coli* genomes. *Microb Ecol* 2010;**60**:708−20.
46. Kaas RS, Friis C, Ussery DW, Aarestrup FM. Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes. *BMC Genomics* 2012;**13**:577.
47. Gordienko EN, Kazanov MD, Gelfand MS. Evolution of pan-genomes of *Escherichia coli Shigella* spp, and *Salmonella enterica*. *J Bacteriol* 2013;**195**:2786−92.
48. Valot B, Guyeux C, Rolland JY, Mazouzi K, Bertrand X, Hocquet D. What it takes to be a *Pseudomonas aeruginosa*? The core genome of the opportunistic pathogen updated. *PLoS One* 2015;**10**:e0126468.
49. Grosso-Becerra M-V, Santos-Medellín C, González-Valdez A, Méndez J-L, Delgado G, Morales-Espinosa R, et al. *Pseudomonas aeruginosa* clinical and environmental isolates constitute a single population with high phenotypic diversity. *BMC Genomics* 2014; **15**:318.
50. Koton Y, Gordon M, Chalifa-Caspi V, Bisharat N. Comparative genomic analysis of clinical and environmental *Vibrio vulnificus* isolates revealed biotype 3 evolutionary relationships. *Front Microbiol* 2015;**5**:803.
51. Polz MF, Alm EJ, Hanage WP. Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends Genet* 2013;**29**:170−5.
52. Donati C, Hiller NL, Tettelin H, Muzzi A, Croucher N, Angiuoli S, et al. Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species. *Genome Biol* 2010;**11**:R107.

Downloaded From Vet-eBooks.com

53. Sanguinetti L, Toti S, Reguzzi V, Bagnoli F, Donati C. A novel computational method identifies intra- and inter-species recombination events in *Staphylococcus aureus* and *Streptococcus pneumoniae*. *PLoS Comput Biol* 2012;**8**:1−11.
54. Smyth DS, Wong A, Robinson DA. Cross-species spread of SCC*mec* IV subtypes in staphylococci. *Infect Genet Evol* 2011;**11**:446−53.
55. Méric G, Miragaia M, de Been M, Yahara K, Pascoe B, Mageiros L, et al. Ecological overlap and horizontal gene transfer in *Staphylococcus aureus* and *Staphylococcus epidermidis*. *Genome Biol Evol* 2015;**7**:1312−28.
56. Bobay LM, Traverse CC, Ochman H. Impermanence of bacterial clones. *Proc Natl Acad Sci USA* 2015;**112**:8893−900.
57. González-González A, Sánchez-Reyes LL, Delgado Sapien G, Eguiarte LE, Souza V. Hierarchical clustering of genetic diversity associated to different levels of mutation and recombination in *Escherichia coli*: a study based on Mexican isolates. *Infect Genet Evol* 2013;**13**:187−97.
58. St Sauver J, Marrs CF, Foxman B, Somsel P, Madera R, Gilsdorf JR. Risk factors for otitis media and carriage of multiple strains o*f Haemophilus influenzae* an*d Streptococcus pneumoniae*. *Emerg Infect Dis* 2000;**6**:622−30.
59. Leung NS, Padgett P, Robinson DA, Brown EL. Prevalence and behavioural risk factors of *Staphylococcus aureus* nasal colonization in community-based injection drug users. *Epidemiol Infect* 2015;**143**:2430−9.
60. Worby CJ, Chang H-H, Hanage WP, Lipsitch M. The distribution of pairwise genetic distances: a tool for investigating disease transmission. *Genetics* 2014;**198**:1395−404.
61. Worby CJ, Lipsitch M, Hanage WP. Within-host bacterial diversity hinders accurate reconstruction of transmission networks from genomic distance data. *PLoS Comput Biol* 2014;**10**:e1003549.
62. Golubchik T, Batty EM, Miller RR, Farr H, Young BC, Larner-Svensson H, et al. Within-host evolution of *Staphylococcus aureus* during asymptomatic carriage. *PloS One* 2013;**8**: e61319.
63. Popovich KJ, Snitkin E, Green SJ, Aroutcheva A, Hayden MK, Hota B, et al. Genomic epidemiology of Usa300 methicillin-resistant *Staphylococcus aureus* in an urban community. *Clin Infect Dis* 2015;**62**:37−44.
64. Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ. Within-host evolution of bacterial pathogens. *Nat Rev Microbiol* 2016;**14**:150−62.
65. Biek R, Pybus OG, Lloyd-Smith JO, Didelot X. Measurably evolving pathogens in the genomic era. *Trends Ecol Evol* 2015;**30**:306−13.
66. Young BC, Golubchik T, Batty EM, Fung R, Larner-Svensson H, Votintseva AA, et al. Evolutionary dynamics of *Staphylococcus aureus* during progression from carriage to disease. *Proc Natl Acad Sci USA* 2012;**109**:4550−5.
67. Didelot X, Nell S, Yang I, Woltemate S, van der Merwe S, Suerbaum S. Genomic evolution and transmission of *Helicobacter pylori* in two South African families. *Proc Natl Acad Sci USA* 2013;**110**:13880−5.
68. Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, et al. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect Dis* 2013;**13**:137−46.
69. Mathers AJ, Stoesser N, Sheppard AE, Pankhurst L, Giess A, Yeh AJ, et al. *Klebsiella pneumoniae* carbapenemase (KPC)-producing *K. pneumoniae* at a single institution: insights into endemicity from whole-genome sequencing. *Antimicrob Agents Chemother* 2015;**59**:1656−63.

Downloaded From Vet-eBooks.com

70. Grad YH, Kirkcaldy RD, Trees D, Dordel J, Harris SR, Goldstein E, et al. Genomic epidemiology of *Neisseria gonorrhoeae* with reduced susceptibility to cefixime in the USA: a retrospective observational study. *Lancet Infect Dis* 2014;**14**:220–6.
71. Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, Brodkin E, et al. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N Engl J Med* 2011; **364**:730–9.
72. Alam MT, Read TD, Petit III RA, Boyle-Vavra S, Miller LG, Eells SJ, et al. Transmission and microevolution of Usa300 MRSA in U.S. households: evidence from whole-genome sequencing. *mBio* 2015;**6**:e00054–15.
73. Frerichs RR, Keim PS, Barrais R, Piarroux R. Nepalese origin of cholera epidemic in Haiti. *Clin Microbiol Infect* 2012;**18**:E158–63.
74. Haley BJ, Choi SY, Grim CJ, Onifade TJ, Cinar HN, Tall BD, et al. Genomic and phenotypic characterization of *Vibrio cholerae* non-O1 isolates from a US Gulf Coast cholera outbreak. *PLoS One* 2014;**9**:e86264.
75. Mutreja A, Kim DW, Thomson NR, Connor TR, Lee JH, Kariuki S, et al. Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature* 2011;**477**: 462–5.
76. Devault AM, Golding GB, Waglechner N, Enk JM, Kuch M, Tien JH, et al. Second-pandemic strain of *Vibrio cholerae* from the Philadelphia cholera outbreak of 1849. *N Engl J Med* 2014;**370**:334–40.
77. Ivers LC, Walton DA. The"first" case of cholera in Haiti: Lessons for global health. *Am J Trop Med Hyg* 2012;**86**:36–8.
78. Hendriksen RS, Price LB, Schupp JM, Gillece JD, Kaas RS, Engelthaler DM, et al. Population genetics of *Vibrio cholerae* from Nepal in 2010: evidence on the origin of the Haitian outbreak. *mBio* 2011;**2**:e00157–11.
79. Katz L, Petkau A, Beaulaurier J, Tyler S, Antonova ES, Turnsek MA, et al. Evolutionary dynamics of *Vibrio cholerae* O1 following a single-source introduction to Haiti. *mBio* 2013;**4**:1–10.
80. Azarian T, Ali A, Johnson JA, Mohr D, Prosperi M, Veras NM, et al. Phylodynamic analysis of clinical and environmental *Vibrio cholerae* isolates from Haiti reveals diversification driven by positive selection. *mBio* 2014;**5**:e01824–14.
81. Alam MT, Weppelmann TA, Weber CD, Johnson JA, Rashid MH, Birch CS, et al. Monitoring water sources for environmental reservoirs of toxigenic *Vibrio cholerae* O1, Haiti. *Emerg Infect Dis* 2014;**20**:356–63.
82. Folster J, Katz L, McCullough A, Parsons M, Knipe K, Sammons S, et al. Multidrug-resistant IncA/C plasmid in *Vibrio cholerae* from Haiti. *Emerg Infect Dis* 2014;**20**:1951–3.
83. Corander J, Marttinen P. Bayesian identification of admixture events using multilocus molecular markers. *Mol Ecol* 2006;**15**:2833–43.
84. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Mol Ecol Notes* 2007;**7**:574–8.
85. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations (DAPC). *BMC Genet* 2010; **11**:94.
86. Lawson D, Hellenthal G, Myers S, Falush D. Inference of population structure using dense haplotype data. *PLoS Genet* 2012;**8**:e1002453.
87. Yahara K, Furuta Y, Oshima K, Yoshida M, Azuma T, Hattori M, et al. Chromosome painting in silico in a bacterial species reveals fine population structure. *Mol Biol Evol* 2013;**30**:1454–64.

Downloaded From Vet-eBooks.com

88. Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, et al. Genomic analysis of diversity population structure virulence and antimicrobial resistance in *Klebsiella pneumoniae* an urgent threat to public health. *Proc Natl Acad Sci USA* 2015;**112**: E3574−81.
89. Ypma RJF, van Ballegooijen WM, Wallinga J. Relating phylogenetic trees to transmission trees of infectious disease outbreaks. *Genetics* 2013;**195**:1055−62.
90. Jombart T, Cori A, Didelot X, Cauchemez S, Fraser C, Ferguson N. Bayesian reconstruction of disease outbreaks by combining epidemiologic and genomic data. *PLoS Comput Biol* 2014;**10**:1−14.
91. Didelot X, Gardy J, Colijn C. Bayesian inference of infectious disease transmission from whole-genome sequence data. *Mol Biol Evol* 2014;**31**:1869−79.
92. Guindon SP, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;**59**:307−21.
93. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;**30**:1312−3.
94. Kozlov AM, Aberer AJ, Stamatakis A. ExaML version 3: a tool for phylogenomic analyses on supercomputers. *Bioinformatics* 2015;**31**:2577−9.
95. Price MN, Dehal PS, Arkin AP. FastTree 2-Approximately maximum-likelihood trees for large alignments. *PLoS One* 2010;**5**:e9490.
96. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 2012;**29**:1969−73.
97. Prosperi M, Veras N, Azarian T, Rathore M, Nolan D, Rand K, et al. Molecular epidemiology of community-associated methicillin-resistant *Staphylococcus aureus* in the genomic era: a cross-sectional study. *Sci Rep* 2013;**3**:1902.
98. McAdam PR, Templeton KE, Edwards GF, Holden MTG, Feil EJ, Aanensen DM, et al. Molecular tracing of the emergence adaptation and transmission of hospital-associated methicillin-resistant *Staphylococcus aureus*. *Proc Natl Acad Sci USA* 2012;**109**: 9107−12.
99. Schierup MH, Hein J. Consequences of recombination on traditional phylogenetic analysis. *Genetics* 2000;**156**:879−91.
100. Hedge J, Wilson DJ. Bacterial phylogenetic reconstruction from whole genomes is robust to recombination but demographic inference is not. *mBio* 2014;**5**:e02158−14.
101. Marttinen P, Hanage WP, Croucher NJ, Connor TR, Harris SR, Bentley SD, et al. Detection of recombination events in bacterial genomes from large population samples. *Nucl Acids Res* 2012;**401**:e6.
102. Didelot X, Wilson DJ. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comp Biol* 2015;**11**:18.
103. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucl Acids Res* 2014;**43**:e15.
104. Shapiro BJ, Friedman J, Cordero OX, Preheim SP, Timberlake SC, Szabó G, et al. Population genomics of early events in the ecological differentiation of bacteria. *Science* 2012; **336**:48−51.
105. Hutter S, Vilella AJ, Rozas J. Genome-wide DNA polymorphism analyses using VariScan. *BMC Bioinforma* 2006;**7**:409.
106. Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. PopGenome: an efficient swiss army knife for population genomic analyses in R. *Mol Biol Evol* 2014;**31**:1929−36.

Downloaded From Vet-eBooks.com

107. O'Neill MB, Mortimer TD, Pepperell CS. Diversity of *Mycobacterium tuberculosis* across evolutionary scales. *PLoS Pathog* 2015;**11**:e1005257.
108. Thomas JC, Godfrey PA, Feldgarden M, Robinson DA. Candidate targets of balancing selection in the genome of *Staphylococcus aureus*. *Mol Biol Evol* 2012;**29**:1175−86.
109. Beaumont MA, Zhang W, Balding DJ. Approximate Bayesian computation in population genetics. *Genetics* 2002;**162**:2025−35.

Downloaded From Vet-eBooks.com

Epidemiology and Evolution of Fungal Pathogens in Plants and Animals

4

P. Gladieux[1], *E.J. Byrnes III*[2], *G. Aguileta*[1], *M. Fisher*[3], *R.B. Billmyre*[2], *J. Heitman*[2], *T. Giraud*[1]

[1]Ecologie Systematique Evolution, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Orsay, France; [2]Duke University Medical Center, Durham, NC, United States; [3]Imperial College London, London, United Kingdom

1. Introduction

The advent of molecular markers offers great tools for studying key processes of parasite biology, such as dispersal, mating systems, host adaptation, and patterns of speciation. Population genetics studies have also valuable practical applications, for instance, for studying the evolution of drug resistance or new virulence. Another reason to study epidemiology and evolution in parasites is that they display a huge diversity of life cycles and lifestyles, thus providing great opportunity for comparative studies to test pathogen-specific questions or general issues about evolution.

About 100,000 species of fungi have been described so far (1.5 million of fungal species are estimated to exist[1]), of which a high percentage obtain nutrients by living in close association with other organisms, mainly plants. Many fungi are" pathogenic and can have important impact on human health or lead to severe economic losses due to infected crops or to animal diseases. Fungal species parasitizing animals and plants are found interspersed with saprophytes and mutualists in fungal phylogenies,[2,3] suggesting that transitions between these life-history strategies have occurred repeatedly within the fungal kingdom.

The two major groups that have been traditionally recognized among the true fungi are the Ascomycota, including the yeasts and filamentous fungi, with several important model species (e.g., *Saccharomyces cerevisiae* and *Neurospora crassa*) and the Basidiomycota, including the conspicuous mushrooms, the rusts, and the smuts. Ascomycota and Basidiomycota have been resolved as sister taxa,[2] and they have been called the Dicaryomycota. Basal to the Dikarya branch several other fungal groups. The Chytridiomycota, for instance, are defined as fungi with flagellated cells and were long thought to be the sister group of all the other true fungi, nonflagellated. Phylogenies since mid-2000s, however, have suggested that the chytrids may in fact be polyphyletic, representing early diverging lineages having retained the ancestral flagellum.[2] Chytrids also encompass plant and animal

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00004-4
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

pathogens. Oomycetes have long been considered as fungi but were recognized during mid-2000s to belong to the distant Stramenopiles.[4] These filamentous organisms, however, share many morphological and physiological characteristics with fungi and continue to be studied by mycologists. We, therefore, also consider oomycetes in this chapter.

Most fungi have been dependent on other organisms for their resources through much of their evolutionary history, in particular fungal pathogens. During the 20th century however, many new fungal diseases have emerged. This is probably due to human activities that have completely modified the ecosystems on earth at a global scale (e.g., climate warming, widespread deforestation, habitat fragmentation and urbanization, changes in agricultural practices, and global trade).[5] Of these, the intensification and globalization of agriculture as well as the increase in international trade and travel have broken down many natural barriers to dispersal causing an unprecedented redistribution of many organisms.[6] Concomitantly, there is growing evidence that these global changes play a key role in the emergence of infectious diseases in humans,[7] wildlife,[8] domestic animals,[9] and plants.[10]

To understand how new diseases emerge, and more generally to understand the spread and maintenance of diseases, it is essential to study dispersal, mating systems, host adaptation, and mechanisms of speciation. The advent of molecular markers offers great tools for studying these key processes of parasite biology.[11,12] The development of full-genome sequencing since mid-2000s, especially among fungi because they have small genomes,[13] has allowed comparative genomics to begin drawing inference on the mechanisms of pathogenicity.[14–20]

In this chapter, we thus describe the main pathogenic fungi, parasitizing humans, animals, and plants, and having important consequences on human health or human activities. We focus on some examples of emerging fungal diseases, on humans, animals, and plants. We then review (1) the modern molecular tools used for epidemiology and population genetics of fungal pathogens, the types of markers most useful, and the different types of analyses that can be performed to unravel their mating systems and dispersal and (2) the advances since 2000s in fungal genomics, in particular the insights that have been gained so far regarding the pathogenic lifestyles.

1.1 Major Human and Animal Pathogenic Fungi

Each of the four major fungal phyla has representatives that cause serious disease in both humans and a vast range of other animals. Although less prevalent than plant pathogens, the animal pathogens pose serious threats to entire animal populations and continue to cause serious morbidity and mortality among immunocompromised patients and otherwise healthy individuals worldwide. In many cases, the incidence of disease is increasing due to a rise in susceptible hosts,

Downloaded From Vet-eBooks.com

while at the same time the treatment options have remained limited in comparison to other classes of pathogens. In the following sections, we summarize the morbidity and mortality associated with several of the major classes of human and animal pathogenic fungi.

1.1.1 *Ascomycetes: The* Candida *Species Complex,* Aspergillus fumigatus, Pneumocystis, *the Dimorphic Fungi, and Others*

Within the fungal kingdom, the ascomycetes harbor the majority of fungal pathogens that afflict humans. Among these, *Candida* species are the most common causes of invasive fungal infections in humans. Infections can range from readily treatable mucocutaneous disorders, although these may be acute in AIDS-infected patients, to severe invasive disease that can result in significant morbidity and mortality, most often occurring in patients with immune system suppression.[21]

Another major cause of human fungal infections is the filamentous pathogen, *Aspergillus fumigatus* and other closely related *Aspergillus* species. Aspergillosis, primarily invasive aspergillosis, is an emerging disease in the immunocompromised population.[22] The spores are widely prevalent in all environments, and are readily inhaled, causing both respiratory and disseminated disease in immunocompromised patients. There is a particularly high incidence of aspergillosis among stem cell and solid organ transplant recipients.[23]

The dimorphic group of pathogenic fungi that cause serious disease in humans in both healthy and immunocompromised individuals includes the pathogens *Histoplasma capsulatum*, *Coccidioides immitis*, *Coccidioides posadasii*, *Paracoccidioides brasiliensis*, and *Penicillium marneffei*.[24–26]

1.1.2 *Basidiomycetes: The Pathogenic* Cryptococcus *Species Complex*

The pathogenic *Cryptococcus* species complex is comprised of seven species previously grouped into *Cryptococcus neoformans* and *Cryptococcus gattii*.[27] They are related basidiomycete yeast species that are common fungal pathogens of both humans and animals. Of the seven species distinguished, *C. neoformans* species are prevalent, ubiquitous worldwide, largely associated with pigeon guano, and a common cause of meningitis in immunocompromised hosts.[28] The *C. gattii* species are generally geographically restricted to tropical and subtropical regions, associated with trees, and commonly infects immunocompetent hosts, although cases in immunocompromised patients also occur.[29] It is estimated that the two clades diverged about 37.5 million years ago, which may explain the observed differences in ecology and host range.[30] Additionally, the "tropical" status of the *C. gattii* species has been challenged by the occurrence of an outbreak of *Cryptococcus deuterogattii* that began in 1999, initially on Vancouver Island, Canada at latitude 49.28°. This emerging

Downloaded From Vet-eBooks.com

infection has since expanded into mainland British Columbia and the Pacific Northwest region of the United States.[31,32]

C. neoformans has been further subdivided into two distinct species, *C. neoformans* and *Cryptococcus deneoformans* (formerly serotypes A and D, respectively) based on unique antigenic profiles and sequence divergence.[27] This distinction is clinically relevant, as serotype A strains cause the vast majority of infections globally, with high incidences in the AIDS and transplant populations.[33] While less prevalent globally, *C. gattii* has also been a significant cause of morbidity and mortality, with high incidences in humans and animals.[34,35] Thus, the *Cryptococcus* species complex remains a global health concern for both humans and a wide range of domestic, agrarian, and wild mammals.

1.1.3 Globally Emerging Fungal Infections in Wildlife Species

While fungi are recognized as serious pathogens to their human hosts, it is also becoming clear that fungal pathogens have the capacity to cause severe disease in wildlife species. For instance, globally spreading chytridiomycosis in amphibians stems from a basal fungal lineage that was never before found to infect vertebrates.[36,37] A new fungal infection of bats called white-nose syndrome has swept through the northeastern United States since 2008, causing the deaths of more than 1 million bats and extirpating some well-known cave roosts.[38] The aetiological agent has been described as an Ascomycete fungus *Geomyces destructans*, related to the human skin-infecting fungus *Geomyces pannorum*.[39]

2. New and Emerging Mycoses

2.1 *Evolution and Emergence of Pathogenic* Cryptococcus gattii *Genotypes in the Pacific Northwest*

As of 1999, *C. gattii* has emerged as a primary pathogen in northwestern North America, including both Canada and the United States.[31,32,40] This outbreak now spans a large geographical range, with levels of infection as high or higher than anywhere else globally, with an annual incidence on Vancouver Island of about 25 cases/million.[41] The only two reports with higher overall levels are one examination of native Aboriginals in the Northern Territory of Australia, and a study conducted in the central province of Papua New Guinea.[41]

The first efforts to elucidate the molecular types of the isolates collected in the Vancouver Island area revealed that two genotypes, now known as VGIIa/major and VGIIb/minor, are responsible for the vast majority of cases.[32] The VGIIa/major genotype was found to be highly virulent in a murine model of infection.[32] In 2007 and 2008, the first reports of *C. gattii* in the Pacific Northwest of the United States were published. The report of Upton and colleagues illustrated the first confirmed case of the Vancouver Island outbreak VGIIa/major in the United States (2006) from a patient in Puget Sound, Washington.[42] Additionally, in 2005, MacDougall and colleagues

Downloaded From Vet-eBooks.com

discovered an increased number of outbreak-related cases on the mainland of British Columbia and related *C. gattii* VGII genotypes in the United States, including one later recognized as a VGIIc/novel isolate. Shortly thereafter, studies by Byrnes et al. documented a large cohort of clinical and veterinary cases from the VGIIa/major outbreak genotype in both Washington and Oregon.[31] These studies also reported VGIIb/minor in the United States, and importantly, defined a novel VGIIc genotype that was unique to Oregon and observed in both human and animal cases.[31]

Phenotypic examinations during mid-2000s have also begun to address several key aspects of the outbreak genotypes. Studies in the mouse model revealed that *C. gattii* isolates from the outbreak induced less protective inflammation than *C. neoformans*, indicating that *C. gattii* may thrive in immunocompetent hosts by evading or suppressing the protective immune responses that normally limit *C. neoformans* disease progression.[43] During mid-2010s it had also been shown that the VGIIc/novel genotype shares similar intracellular proliferation rates, mitochondrial morphology, and murine virulence characteristics with the VGIIa/major genotype, further supporting the hypothesis that the genotypes seen in the region are uncharacteristically enhanced for virulence.[44]

Further phylogenetic analysis within the *C. deuterogattii* species has identified South America as a likely ancestral origin of the species.[27] In addition, three whole-genome resequencing studies were performed that provide evidence that the three clonal groups in the Pacific Northwest outbreak were introduced from distinct proximal locations, with VGIIa likely coming from South America and VGIIb from Australia, while VGIIc still has no known proximal origin and may therefore either have arisen locally or be from an undersampled environmental niche in another locale.[45–47]

Since 2000s, we witnessed the emergence and expansion of a tropical/subtropical pathogen into a temperate climate, leading to the formation of a multidisciplinary *C. gattii* working group that was established to address the epidemiology, clinical features, and basic science questions surrounding this outbreak.[48] Substantial progress has been achieved in addressing the molecular epidemiology and expansion of the outbreak, and also the phenotypic characteristics that make these genotypes unique.

2.1.1 *The Global Emergence of the Amphibian Pathogen* Batrachochytrium dendrobatidis

The ability of fungi to cause severe disease in nonhuman vertebrate species has been dramatically illustrated by global declines in amphibian biodiversity caused by the fungus *Batrachochytrium dendrobatidis* (*Bd*). Only discovered in 1997[49] and named in 1999,[50] *Bd* is a basal fungal lineage in the Chytridiomycota.[2] Many species of chytrid have been described in aquatic environments and soils, as free-living or commensal organisms, and as parasites of algae, invertebrates, and fungi.[51] Of these, *Bd* is unique in that it is the only chytrid known to parasitize vertebrates.[52] *Bd* is now known to be widespread in all continents except Antarctica (where amphibian hosts do not occur). A global-mapping project for this pathogen has shown that *Bd* infects over 350 species

Downloaded From Vet-eBooks.com

of amphibian, and has been implicated in driving the declines and extinctions of over 200 of these (http://www.spatialepidemiology.net/bd-maps/).[37,53]

Following the discovery that *Bd* was a driver of declines in amphibian species in Australia, the Americas, and Europe, much attention has been focused on finding out how *Bd* was being spread, and from where. In eastern Australia, prospective and retrospective sampling of amphibians has shown that populations were initially *Bd*-negative prior to 1978 followed by an expansion north and south from a center in southern Queensland; western Australia was *Bd*-negative until mid-1985 whereupon the spread of disease was detected and documented. Mesoamerica has witnessed a rapid wave-like front of expansion from an apparent origin in Monteverde, Costa Rica, southward at estimated rates of between 17 and 43 km/year, and during late 2000s jumped the Panama Canal.[54] The epidemic front of chytridiomycosis along the North−South transect of Central America has been predictable to the extent that researchers have been able to anticipate the arrival of *Bd* in uninfected regions, and to document the collapse of the amphibian community upon arrival of the pathogen and the onset of chytridiomycosis.[55]

Given these patterns of declines, where is the original source of *Bd*? Answers to this question have been sought by attempting to identify geographical regions where *Bd* has had a long and stable association with host species, indicative of coevolution, as well as substantially increased levels of genetic diversity when compared against the various regional epizootics. One such study[56] has identified Africa as a potential source of the panzootic. Histology on historical museum specimens showed that *Bd* has infected amphibians in Southern Africa since at least 1938, and the "*Bd* Out of Africa" hypothesis was coined to suggest that *Bd* was spread around the world via the extensive trade in the African clawed frog *Xenopus laevis* since 1930s onward. Recent discovery that genotypes of *Bd* occur in the Japanese archipelago that appear basal to the panzootic lineage suggests that there may yet be other potential sources of *Bd* diversity.[57] Therefore, the overarching question on the origin of *Bd* remains unanswered to date. What is clear, however, is that the global trade in amphibians is a potent force in spreading *Bd* into naive populations and species.

Importantly, all population genetic studies so far have shown that *Bd* exhibits levels of heterozygosity that are consistent with a predominately asexual mode of reproduction. Levels of heterozygosity were not uniformly distributed across the genome, but were significantly reduced on the largest inferred chromosome where loss of heterozygosity (LOH) had occurred. This pattern of LOH is not consistent with sexual reproduction and segregation, but rather with a model of chromosome-specific variation in mitotic (somatic) recombination, a process that is well documented in other fungi including the diploid pathogenic fungus *Candida albicans* that exhibits vegetative diploidy.[58]

Despite the apparent rapid spread of *Bd* and the high degree of relatedness between isolates, data are accumulating showing that genotypes differ significantly in their virulence. A 2010 study[59] showed that the sporangia of five isolates of *Bd* from the Balearic Islands of Mallorca, all with identical genotypes, were similar in size, but differed significantly from those isolates recovered from amphibians in mainland Spain and the UK. When the virulence of a Mallorcan isolate of *Bd* (TF5a1) and a

Downloaded From Vet-eBooks.com

UK isolate of *Bd* (UKTvB) was assayed in *Bufo bufo*,[59] the Mallorcan strain of *Bd* was avirulent in comparison against the UK strain of the pathogen.

2.2 *Origin of Human Pathogens:* Cryptococcus *and* Candida *From Saprobes Associated With Insects*

The origin and evolution of pathogens remain central questions in studies of both plant and animal diseases. One method to examine the likely origins of pathogens is to phylogenetically place the species into the context of closely related saprobic relatives. As mentioned earlier *Cryptococcus* and *Candida* represent major classes of mammalian fungal pathogens, and in both cases their closest related species are associated with insects. Although these sibling species are often less studied than their medically relevant counterparts, they offer important insights into the evolution of the animal pathogens and how these pathogenic species might have arisen from insect-associated saprophytes.

Phylogenetic analyses indicate that the *Cryptococcus* species complex likely arose from the *Tremella* lineage and that it clusters closely with the Tremellales, Trichosporonales, Filobasidiales, and the Cystofilobasidiales.[60–62] Several of the species within these lineages are saprophytes that are commonly associated with insect debris, leading to the hypothesis that the pathogens emerged from an association within this environmental niche.[63] In support of this hypothesis, *C. gattii* has been isolated from both insect frass and wasp nests and *C. neoformans* has been isolated from honeybee hives, indicating that these animal pathogens may still in some cases act as an insect-associated saprophyte in the environment.[61]

3. Plant Pathogenic Fungi

Although several important fungal pathogens attack animals, land plants have probably been the main nutrient source of fungi through much of their evolutionary history, given the predominance of plant saprophytes, pathogens and mycorrhizal species in fungi.[2,3] Collectively, fungi cause more plant diseases than any other group of plant pest (such as viruses or bacteria) with over 8000 species shown to cause disease. Plant diseases caused by fungi exhibit a huge diversity of symptoms. Pathogenic fungi can indeed be responsible for example for lesions on leaves or on flowers, for stem cankers, for root and fruit rot, or can sterilize plants.

Fungal pathogens are therefore a serious concern for agriculture, as they reduce crop yield and lower product quality by attacking cultivated plants and their products (seeds, fruits, and grains). Nearly all crops have their pathogenic fungi, and often several of them, from cereals to corn, rice, potatoes, beans, peas, soybean, fruit trees (including coffee and cacao), and ornamental plants and trees. Some of the world's great famines and human suffering can be blamed on plant pathogenic fungi. In the United States alone, hundreds of millions of bushels of wheat have been lost in epidemic years to stem rust (*Puccinia graminis tritici*). Rice blast, caused by the

Downloaded From Vet-eBooks.com

fungus *Magnaporthe oryzae*, is an important disease on rice, among many other diseases. It is found wherever rice is grown, it is always important, and it is always a threat. The *Botrytis* gray mold is a common disease of greenhouse floral crops and all ornamental plants can be infected by powdery mildews. These are only few examples of the many pathogenic fungi devastating crops.

Pathogenic fungi are also widespread in natural ecosystems, with great impacts on the compositions of natural communities. Forest trees for instance are attacked by many pathogenic fungi. *Armillaria* root disease, causing branch dieback and crown thinning, is often one of the most important diseases of trees in temperate regions of the world, especially in native forests. The most infamous tree diseases include Dutch elm disease, caused by *Ophiostoma* species, chesnut blight, caused by *Cryphonectria parasitica*, and Sudden oak death, ramorum leaf blight, ramorum shoot blight, all caused by the oomycete *Phytophthora ramorum*. These diseases have dramatic consequences on forest composition and their associated biota, with some tree species even disappearing from continents. For instance, the chestnut blight fungus caused the death of 80% of the native American chestnut trees throughout eastern forests from Maine to Georgia during the first half of the 20th century.

4. New and Emerging Plant Diseases

Fungi are also responsible for about 30% of emerging diseases in plants (sensu lato, i.e., including Oomycetes), which is three times more than for emerging diseases in humans or wildlife.[10] These patterns of fungal disease emergence in plants have elicited great concern for several reasons. First, epidemics caused by invasive pathogens have been repeatedly reported to alter natural ecosystems.[10,36,64] Well-documented examples of emergent diseases in natural plant communities include some of the ones mentioned earlier, such as the spread of *C. parasitica* that eliminated the dominant chestnut forests throughout eastern North America at the end of the 19th century. *Phytophthora cinnamomi* that threatens native forests throughout Australia is also an emerging disease.[10,64] Such dramatic diseases not only affect the host plants, but also the whole associated fauna, including insects, birds, and mammals.

Second, our primary food production is at risk due to emerging crop diseases; the most dramatic historical example being the Irish Potato Famine caused by *P. infestans* on cultivated potato in the beginning of the 20th century.[65] Other examples of invasive fungi parasitizing crops include *Plasmopara viticola*, an oomycete causing the grapevine downy mildew that has been introduced from North America to Europe during 19th and 20th centuries. Crop plants are in fact particularly susceptible to the emergence of new diseases because of the large-scale planting of genetically uniform varieties.

There has been an increasing focus on identifying the factors that drive the emergence of new fungal diseases.[10,64,66,67] As mentioned earlier, introduction of pathogens in a new area is one of the most obvious causes. It has been estimated that

Downloaded From Vet-eBooks.com

between 65% and 85% of plant pathogens worldwide are alien in the location where they were recorded.[68]

5. Modern Molecular Epidemiological Tools for Investigating Fungal Diseases

To understand the dynamics of fungal diseases and the dynamics of emergence of new diseases, epidemiology is a necessary step. Epidemiology is indeed a discipline concerned with understanding the factors affecting the dynamics of disease in space and in time, with an emphasis on being quantitative and predictive. During the 2000s, the integration of molecular biology into traditional epidemiological research has revolutionized the discipline.[69,70] This led to the development of a new field, *molecular epidemiology*, which addresses epidemiological problems using "the various molecular methods that aim to identify the relevant units of analysis of pathogens involved in transmissible disease."[71] Two methods in particular are now predominant in molecular epidemiological studies of fungal pathogens: MLST, for Multi-Locus Strain Typing[72] and MLMT, for Multi-Locus Microsatellite Typing.[73] Currently, new advancement in next-generation high-throughput sequencing techniques mean that MLST and MLMT typing schemes are on the brink of being absorbed into whole-genome single-nucleotide polymorphism−typing platforms. For instance, a 2009 MLST study of the chytrid fungus *B. dendrobatidis*,[74] suggested as a principal cause for the worldwide decline of amphibians, found the global epidemic owes to the global dispersal of a single genotype. These data were used to argue that the observed low allelic diversity and high heterozygosity provide strong support that the fungus is a novel pathogen introduced into naive host populations, over the alternative hypothesis that the species is an endemic pathogen whose emergence is due to changes in the environment. By contrast, an MLST study of *C. immitis*, the etiological agent of coccidiomycosis revealed that the epidemic observed in California during the early 1990s was not due to the emergence of a virulent genotype but rather governed by the synchrony of environmental factors.[75]

MLMT-based techniques are more useful in discriminating genotypes within species and inbred populations, than among species, which make their use complementary to MLST. A 2004 study[76] sequenced three housekeeping genes in outbreak isolates of *P. ramorum*, the aetiological agent of the devastating "Sudden Oak Death" disease. This study showed that all sequences were identical among all isolates, and therefore completely uninformative on the nature of epidemic. MLMT tools proved very useful in tracking the pathogen as it spread in the United States.[77] Analyses of MLMT data provided evidence of a historical link between nursery and wild populations of the pathogen, and identified three common genotypes as the likely founders of the Californian epidemics.[78] MLMT also provides a useful tool to infer the source and type of primary pathogen inoculum, which are often impossible to identify by direct observation or using the traditional epidemiological approach of studying the distribution of disease foci.[79]

Downloaded From Vet-eBooks.com

6. Population Genetics of Pathogenic Fungi

Population genetics is also needed to understand fungal diseases. By providing an understanding of the processes that shaped the structure of a pathogen species in the past, population genetics offers the opportunity to forecast the emergence of genotypes, populations, or species with detrimental characteristics for human affairs,[80,81] and also to inform practical attempts to bring fungal pathogens into durably effective human control.[82]

6.1 Reproductive System

Fungi present a striking diversity of life cycles, and studying their reproductive biology is a challenging task. Yet, this information is critical to assess the risk posed by pathogens and for the design of disease management strategies.[80] For instance, outcrossing promotes genetic exchange and can hence accelerate the spread of new mutations in combination with other beneficial alleles, which is critical in the context of an arms race between hosts (or the humans that breed or grow them) and pathogens. By contrast, selfing or asexual reproduction provides insurance of reproduction for species having a low probability of finding a mate, and these species can therefore invade distant territories more easily and/or more rapidly.[83] Asexual reproduction is also an expeditious way of multiplying rapidly favorable combinations of genes built by past selection,[84] and a more efficient strategy of transmitting genes to the next generation.

6.1.1 Analysis of the Reproductive System

The identification of populations and species is an essential prerequisite to the study of the reproductive mode and mating system. Hidden population subdivision or cryptic species within the units defined to perform analyses can indeed lead to erroneous conclusions on the reproductive biology of a fungus. This causes deviations from random mating or random association among alleles. A well-known example is the Wahlund effect, where the failure to detect population subdivision influences measures of inbreeding and association among alleles at different loci and leads to the same signal as inbreeding.

The most immediate consequence of asexual reproduction is the occurrence of repeated identical genotypes. The ratio of the number of multilocus genotypes found over the sample size can give an idea as to the rate of asexual reproduction, ranging from zero for a completely clonal population to one for a sexually reproducing population. Many populations of plant pathogens actually fall between the two extremes, having annual sexual cycles and asexual epidemic phases that amplify clones.[85]

Under random mating, the frequency of multilocus genotypes is expected to be equal to the product of the allelic frequencies. Deviation from this expectation (or linkage disequilibrium) can hence serve as a test for random mating. A first approach is to analyze linkage disequilibrium between pairs of loci. The lack of association among pairs of loci in two isolated groups of the agent responsible for gray mold (*Botrytis*

Downloaded From Vet-eBooks.com

cinerea), for example, supported regular events recombination despite the absence of a sexual structure in field observations.[86] Another, more powerful, approach is to analyze linkage disequilibrium over multiple loci. This forms the basis of the test based on the index of association I_A.[87] The I_A statistic relies on the variance of the number of differences among individual allelic profiles. This variance is higher than expected if mating is nonrandom due to an excess of very close and very large distances among individuals. This procedure has been applied to investigate the reproductive mode of *P. marneffei*, the causal agent of biverticilliate mycosis in mammals. Analyses revealed very high and significant values of the index of association statistic,[88] providing one of the very rare cases of a fungus showing no evidence of recombination by population genetic criteria.[83] There are also several examples where the index of association suggested the existence of cryptic sexual reproduction in fungal pathogens in species where sex has not been observed in nature, such as the human pathogens *C. immitis*.[89]

In diploids or dikaryotic fungi, insights into the reproductive mode can be provided by the use of Wright's *F*-statistics F_{IS}, a measure of the deviation from random mating. For instance, the finding of F_{IS} values nonsignificantly different from zero allowed[90] to conclude to the existence of sexuality in Chinese populations of *Puccinia striiformis* f. sp. *tritici*, a fungus showing a highly clonal population structure in other regions of the world. Another application of this approach[74] revealed a significant excess of heterozygous genotypes for half of the loci surveyed (i.e., $F_{IS} < 0$) in worldwide samples of the amphibian-killing fungus *B. dendrobatidis*, suggesting a predominantly asexual mode of reproduction.

A number of methods have also been developed to estimate the population recombination rate (ρ) from haplotype data representing multiple positions in the genome (i.e., typically, moderate to large genomic dataset).[91]

6.2 Dispersal, Migration, and Gene Flow

Dispersal is the movement of gametes or individuals. Parameters of dispersal can be estimated by (1) direct methods, relying on direct observation of dispersing individuals at particular life-history stages, which provides a measure of actual dispersal or (2) by indirect methods that use the changes in some characteristics of populations caused by movement of individuals, and provide a measure of effective dispersal.[92,93] Because the movement of individuals obviously leads to movement of genes, the study of dispersal is tightly related to the study of gene flow (direct methods) and the monitoring of particular genotypes (indirect methods). The two types of methods are treated together here.

For fungal pathogens, in practical terms, some of the most unfortunate consequences of gene flow for human affairs include immigration of genotypes capable of defeating a resistance gene, exchanges of alleles allowing resistance to antifungal molecules (and more generally the spread of variants with increased pathogenicity), increase in population size which in turn increases the probability of accumulating mutations and increase the efficacy of selection (and the possibility of selective sweeps). The degree of gene flow is also of central importance in the formation and maintenance

Downloaded From Vet-eBooks.com

of pathogen species. Humans have moved many pathogens far beyond their natural dispersal limits, and it is a safe bet that many pathogens are still transported among continents today.[80,94] These introductions likely have set the stage for the formation of reproductively isolated populations adapting to local hosts or environments[81] or for secondary contacts followed by introgression or hybridization among species.[66] Gene flow is thus a critical target for disease management tactics.

6.2.1 Rate and Direction of Gene Flow

Pathogenic fungal species are often organized into discrete populations. Population genetics usually assumes a simple model of n populations, each of which is equally likely to receive and give migrants to and from each of the other populations. Under this model, providing additional simplifying assumptions, a relationship between $N_e m_e$ (N_e being the effective size of each population; m_e being the effective migration rate between populations) and F_{ST} (an F-statistic that measures of genetic differentiation among populations by quantifying the differences in allele frequencies between populations) can be derived: $F_{ST} \approx 1/(1 + 4\, N_e m_e)$. This approach has been severely criticized by some authors[95,96] who raised concerns about the unrealistic assumptions under the n-island model (constant population sizes, symmetrical migration at constant rates, no selection, and persistence for periods of time long enough to achieve migration−drift equilibrium). Even though they do not provide reliable estimates of rates of gene flow, measures of population differentiation can nonetheless be used to gain information on the history of dispersal. Several studies reported very low differentiation among samples of fungal pathogens of agricultural crops or forestry trees from different localities across a continent (e.g., *Venturia inaequalis*,[97,98] *Melampsora larici-populina*[99]).

The coalescent theory[100] relates patterns of common ancestry within a set of genes to the structure of the populations from which they were sampled. In coalescent models, patterns of relationships among genes are represented by a genealogy, and the structure of the population is represented by parameters such as population size, rates of population growth, or—what is relevant to the present discussion—rates and directions of gene flow. Both the genealogy and the parameters are generally unknown, and the one usually wants to estimate the parameters of the model. It is generally impossible to jointly consider all possible ancestral relationships and parameter values and to search for the combinations that maximize the probability of the model. Instead, approaches have been developed that simultaneously explore many relatively probable genealogies (loosely speaking, irrelevant genealogies are disregarded) and parameter values (see Refs. 101,102 for reviews). These approaches are collectively referred to as "coalescent genealogy samplers." Several methods relying on coalescent genealogy samplers were designed to estimate, among other parameters, rates of gene flow between species or populations.[103,104] These methods offer the advantage of allowing less restrictive models than the more traditional methods presented earlier. These methods have been successfully applied to infer the ancestral routes of colonization for several fungal globally distributed plant pathogens such as the barley scald pathogen *Rhynchosporium secalis*,[105] and the apple scab pathogen *V. inaequalis*.[97]

Downloaded From Vet-eBooks.com

Methods based on coalescent genealogy samplers remain computationally demanding. For many datasets and models of population structure, they even remain computationally intractable. As a result, there is an increasing interest in developing alternative approaches that are faster and easier to implement. One of the most promising approaches is approximate Bayesian computation[106]; it has been shown to be particularly powerful to determine the origin and routes of introduction of invading pest species,[107–109] and it is very likely that it will also provide important insights into the history of fungal pathogens.

6.2.2 Dispersal Distance

There is a considerable interest in estimating the distance fungal pathogens disperse at agriculturally relevant scales, such as fields or production areas. This information can be inferred from patterns of genetic variation by fitting a model of isolation by distance. The slope of the regression of differentiation statistics (e.g., F_{ST}) onto the log-transformed geographical distance among individuals or populations allows estimation of the product of D, the population density, and σ^2, the second moment of dispersal distance.[110] For fungal pathogens that alternate asexual and sexual reproduction during their life cycle, these methods are not suitable due to the occurrence of repeated genotypes.[111] Variograms (i.e., plots of the semivariance in number of differences between genotypes against distance) are efficient tools to estimate the degree and extent of spatial genetic structure accounting for autocorrelation (which is the tendency that nearby observations to be more similar than distant ones). Variograms were used to study dispersal in the chestnut blight fungus (*C. parasitica*), showing that asexual spores probably disperse over several hundred meters, which is a far larger spatial scale than previously thought.[111]

6.2.3 Distribution of Gene Flow in Time and Along the Genome

The coalescent-based implementation of the isolation-with-migration model in the IM and IMa program[103,112] offers the opportunity to gain valuable insights into the history of gene flow between species. An interesting feature of the program is that counts and dates of migration events in sampled genealogies can be recorded during the course of the MCMC at stationarity for each locus, to obtain the migration time distribution. IM was used to demonstrate that the wheat pathogen *Mycosphaerella graminicola* emerged in the fertile crescent at the time of wheat domestication following a series of introgressions from populations infecting three different uncultivated grasses.[113] Estimates of the time of gene flow events indicated that populations from wheat and uncultivated grasses diverged in the face of gene flow but are now genetically isolated.

6.3 Population Subdivision

Fungal pathogens, like all organisms, are not homogeneously distributed across the environment, which can lead to genetic structure. There are two main sources of population subdivision in fungal pathogens: geography and hosts. While some species

Downloaded From Vet-eBooks.com

have very broad host ranges (e.g., the amphibian pathogen *B. dendrobatidis*, >350 host species[53]; or the gray mold *B. cinerea*, >235 host species[114]), others display clear subdivisions that correspond to the host of origin of populations (e.g., *V. inaequalis*[115]). Such host-specific divergence may evolve as a consequence of limited dispersal or of trade-offs in adaptation. Among pathogen species found on a single host, some species display clear geographically distinct populations (e.g., the mammalian pathogen *H. capsulatum*[116] or the white campion smut *Microbotryum lychnidis-dioicae*[117]), while others appear to have global distributions such as the human pathogen *A. fumigatus*.[118] These patterns of geographical subdivision result from a complex interplay between contemporary and historical gene flow processes.

Understanding the origin of population subdivision is fundamental to our knowledge of the mechanisms responsible both for disease emergence and for the biodiversity of fungi. Four main approaches are available to analyze population subdivision: measures of differentiation, evolutionary trees, multivariate methods, and model-based clustering algorithms.

6.3.1 Measures of Differentiation

Population subdivision can be assessed by calculating differentiation indices (e.g., F_{ST}) between pairs of populations. The AMOVA framework summarizes population differentiation into *F*-statistics by partitioning molecular variance among the different hierarchically nested levels of sampling represented in a dataset (which can be localities, host species, regions, continents, and so on). The main drawback of this procedure is that the sampling units must be assigned into given hierarchical subdivisions by investigators, which may be a relevant issue.

6.3.2 Evolutionary Trees

The most traditional approach to track population subdivision from genetic data is to build an evolutionary tree. Two main classes of evolutionary trees construction methods are available: (1) clustering methods use an iterative method (e.g., neighbor joining) to combine samples in a hierarchical fashion; (2) searching methods consider a range of possible trees and choose the ones that best fit the data according to an optimality criterion (such as maximum parsimony, maximum likelihood, or maximum Bayesian probability).[119]

Evolutionary trees are appealing because they provide a graphical representation of the relationships among samples.[120] When constructed from multilocus data, evolutionary trees can be very useful for exploratory data analysis or for visualizing the main subdivisions within a dataset. When interpreting an evolutionary tree, there are two main reasons to be cautious: (1) the stochastic variance in evolutionary trees (the problem being greater for evolutionary trees based on a single locus) and (2) the inadequacy of a bifurcating model when applied at the intraspecific level. The stochastic variance in evolutionary trees is due to the fact that different loci that have passed through the same demographic history, leading to evolutionary trees that vary widely in topology and branch lengths.[120] The other potential issue is that

Downloaded From Vet-eBooks.com

bifurcating models may not be appropriate to represent relationships at the intraspecific level. An alternative to tree-based approaches for representing relationships among samples is to use a network. Several methods of network reconstruction have been developed. Networks offer the advantage over evolutionary trees of being able to incorporate persistent of ancestral nodes, multifurcations, and reticulations.[121]

6.3.3 Model-Based Bayesian Clustering Algorithms

The aim of model-based Bayesian clustering algorithms (or assignment methods) is to infer groups of individuals (called clusters or populations) that "fit some genetic criteria that define them as distinct groups."[122] The use of a clustering method is an almost unavoidable step in every population genetic study. This field has been flourishing for a decade, and we do not give an extensive description of all the methods currently available. The most popular program is STRUCTURE.[123] The method assumes a model in which there are K clusters, each of which being characterized by a set of allele frequencies. Assuming Hardy−Weinberg and linkage equilibrium within clusters, the program simultaneously estimates allele frequencies in each cluster and then assigns every individual probabilistically to clusters.

6.3.4 Multivariate Methods

The principle of multivariate analyses, when applied to genetic variation among individuals or populations, is to extract and summarize multivariate genetic information into a few synthetic variables.[124] Methods, such as principal component analysis, have been very early applied to population genetics questions.[125] Multivariate methods offer three main advantages. A first advantage is that they perform much faster than methods that are based on evolutionary trees or Bayesian clustering algorithms. A second advantage is that these methods make no assumption of population structure, such as Hardy−Weinberg or linkage equilibrium. This can be particularly useful for fungal pathogens with asexual or partially asexual modes of reproduction, for which Bayesian clustering algorithms present a high risk of producing spurious assignments.[123] A principal component analysis was applied to investigate the origin of French populations of the chestnut blight fungus, a species in which high rates of asexual reproduction and maybe also of intra-haploid sexual reproduction (allowed by homothallism) result in high frequencies of repeated multilocus genotypes.[126] Analyses revealed three distinct genetic lineages with separate geographical distributions, suggesting independent introduction events with limited gene flow among lineages descending from the three original groups of founding strains.

6.4 Conclusion

Empirical population genetics studies have revolutionized our understanding of fungal pathogen evolutionary biology. The distribution range of pathogens (in space, and on hosts), their reproductive system and transmission pathways are crucial features of pathogen biology that would have remained inaccessible based solely on phenotypic

Downloaded From Vet-eBooks.com

data and without the powerful inferential framework of population genetics. How could have we showed that "everything is not everywhere" and that many broadly distributed fungal pathogens are actually subdivided into populations constrained to small geographical areas? How could have we known that only very few fungal pathogens are ancient strictly asexual species and that the deuteromycota do not constitute a formal phylum of fungi? The upcoming flood of genomic data should galvanize investigations on central topics such as the evolution of reproductive systems,[127,128] the acquisition of virulence to new hosts, resistance to disease control strategies, and the evolution of reproductive isolation.[83,129,130]

7. Genomics of Fungi: What Makes a Fungus Pathogenic?

7.1 Comparative Genomics of Plant Pathogens

In this section we are interested in exploring the genomic characteristics that allow some fungi to infect plants and animals.[14,131–133] The pathogenic fungi are most often opportunistic.[134,135] Their capacity to derive nutrients from a large range of plant hosts appears to rely on a battery of genomic resources that are the result of different evolutionary processes. Perhaps the most important source of new genes and gene functions that are specific of fungal pathogens are derived via expansions of gene families that facilitate the infection of the host.[136–138] Typically, these gene families include cell surface receptors such as the G-protein-coupled receptors, which bind exogenous ligands and participate in signaling cascades[139]; secreted proteins, which constitute a diverse group of small peptides such as toxins, proteinaceous effectors, and hydrolytic and degrading enzymes[140]; protein effectors that suppress plant defenses and alter cellular metabolism[141]; and secondary metabolites such as nonspecific and host-specific toxins.[142] Gene families typically expand by gene duplication, which in fungal genomes range from whole-genome duplications[143–145] to several instances of tandem duplications, such as events involving pathogenicity-related gene families including adhesins,[146,147] the ABC transporters, and major facilitator superfamily (MFS) drug efflux systems that help fungi detoxify products from the plants defenses,[148] the multidrug-resistance transporter families,[149] and major surface glycoproteins. Gene duplications related to adaptations to the pathogenic lifestyle have also been documented, as in the case of the oxidative phosphorylation pathway, whose components have evolved by functional divergence with several instances of gene loss and duplication.[141,150,151] Following duplication, rapid rates of evolution and positive selection can give rise to novel gene functions that allow the fungus to coevolve with its host or to infect new hosts. In fungal genomes, positive selection has been found to act in the evolution of functionally important gene families, in particular those that confer an adaptation to a pathogenic lifestyle.[152] These include genes coding for defense systems or for evading host-resistance mechanisms, toxic protein genes, and other virulence-related genes.[153] Positive selection in the plant defense R-genes is frequently followed by coevolution in the avirulence genes of the fungal parasite.[154] This gene-for-gene

Downloaded From Vet-eBooks.com

interaction with corresponding responses in both the host and the parasite genomes is referred to as an "arms-race" process.

In terms of the structure of fungal genomes, it has been shown that genes encoding biochemical products aiding in infection are often clustered together.[155] Clustering of important gene families appears to offer several advantages for pathogenicity.[118,156] Indeed, evidence shows that fungal genes interacting in the same metabolic pathway tend to be clustered together.[157] Transposable elements are another class of genomic elements that have also been shown to play a significant role in enhancing the pathogenic capacities of fungi. In several pathogenic fungi, including *Leptosphaeria maculans* and *Magnaporthe grisea*, sequences coding for avirulence genes are found in genomic regions dense with transposable elements,[158–161] potentially contributing to the extreme variability of avirulence genes that are associated with host–pathogen coevolution. Telomeres are rapidly evolving genomic regions particularly prone to the accumulation of transposable elements, and they sometimes contain avirulence genes, thereby playing a role in host adaptation.[162,163] Sometimes, the genes that confer pathogenicity to fungi come from other species, either via horizontal gene transfer (HGT) or hybridization. Although HGT is not as pervasive in fungal genomes as it is in bacteria, it appears to have occurred multiple independent times and providing adaptation.[132,164–170] Occasionally, complete clusters are speculated to have been horizontally transferred.[171] Finally, hybridization is another way to mix genes and produce new crosses with increased pathogenic capacities.[172]

Fungal genomes are extremely plastic. This is highlighted by the different genomic processes that have generated a versatile repertoire of biochemical functions that allow fungi to colonize a diverse range of environments and also to establish relationships with other species, either by infection or by symbiosis, with an extensive array of partners. New genomic data will continue to fascinate us with examples of amazing potentials for adaptation.

7.2 Comparing Animal and Plant Pathogens

Pathogenic fungi are mostly intracellular pathogens, indicating that at some point during the interaction between the host and the invading species the pathogen lives inside the host cell. Despite the variety of intracellular fungal pathogens infecting both plant and animal cells in seemingly unique ways, there are only few general solutions to the challenge of penetrating and surviving inside host cells.[173] Indeed, the problem represented by intracellular infection has been tackled by convergent solutions that have evolved in parallel in the different fungal lineages[174] of both plant and animal pathogens. It is interesting to note that among fungi there appears to be many more species that parasitize plants than animals.[175] The reasons for this imbalance are not very clear and deserve further attention.

Interesting reviews highlighting similarities and contrasts between animal and plant fungal pathogens are available.[176–178] The genomes of fungal animal pathogens have not been as extensively studied as phytopathogens. More research needs to be conducted and more animal pathogens need to be sequenced before we have a comprehensive view of the genetic basis, if any, of the differences between the fungal genomes of

Downloaded From Vet-eBooks.com

plant and animal pathogens. Most mechanisms and gene functions may be shared, as has been shown by a study of the NLP toxin whose fold is conserved and shows similarities with that of bacteria,[179] hence we can speculate about lineage- and host-specific genes and gene functions in each case.

8. Conclusion

Comparative genomic studies in plant pathogenic and symbiotic fungi, although still in the early stages and limited to a few pathogens, have already brought many insights into the evolution of the pathogenic lifestyle, in particular into the mechanisms of virulence and host adaptations. There is a marked bias in the sequencing efforts toward pathogenic fungi, but current projects are covering the fungal genomes of species with very diverse lifestyles, that will hopefully allow us to gain further insights into the genomics of pathogenicity.

Regarding epidemiology, molecular methods have much to offer to the study of fungal pathogens, allowing elucidation of ecological and microevolutionary processes. Population genetic approaches have provided important insights for some fungal pathogens on their mating systems, dispersal, and population structure. However, much wider employment of these methods is warranted to study fungal pathogens, where it is still too restricted, although much progress has been made since 1990s. Microsatellite markers in particular are very powerful tools[180] and should be more widely used for population studies in fungi, despite the technical challenges of their isolation in this Kingdom.[181] Further, new methods to analyze data are being developed at a rapid pace, using for instance the Bayesian or the coalescence frameworks, or coupling geography and genetics to unravel migration and speciation histories, which should allow even more powerful inferences on the evolutionary processes. However, further theoretical development is badly needed to apply the extant molecular methods to the variety and specificities of the fungal life cycles, such as pervasive clonality and alternation between haplo- and diploid phases.[182,183]

References

1. Hawksworth DL. The fungal dimension of biodiversity: magnitude, significance, and conservation. *Mycol Res* 1991;**95**:641–55.
2. James T, et al. Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* 2006;**443**:818–22.
3. Berbee ML. The phylogeny of plant and animal pathogens in the Ascomycota. *Physiol Mol Plant Pathol* 2001;**59**:165–87.
4. Keeling PJ, Burger G, Durnford DG, Lang BF, Lee RW, Pearlman RE, et al. The tree of eukaryotes. *Trends Ecol Evol* 2005;**20**(12):670–6.
5. Kareiva P, Watts S, McDonald R, Boucher T. Domesticated nature: shaping landscapes and ecosystems for human welfare. *Science* 2007;**316**(5833):1866–9.

Downloaded From Vet-eBooks.com

6. Kolar CS, Lodge DM. Progress in invasion biology: predicting invaders. *Trends Ecol Evol* 2001;**16**(4):199−204.
7. Tatem AJ, Rogers DJ, Hay SI. Global transport networks and infectious disease spread. *Adv Parasitol* 2006;**62**:293−343.
8. Daszak P, Cunningham AA, Hyatt AD. Emerging infectious diseases of wildlife − threats to biodiversity and human health. *Science* 2000;**287**:443−9.
9. Cleaveland S, Laurenson MK, Taylor LH. Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. *Philos Trans R Soc Lond Ser B-Biol Sci* 2001;**356**(1411):991−9.
10. Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P. Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends Ecol Evol* 2004;**19**:535−44.
11. Criscione CD, Poulin R, Blouin MS. Molecular ecology of parasites: elucidating ecological and microevolutionary processes. *Mol Ecol* 2005;**14**(8):2247−57.
12. Giraud T, Enjalbert J, Fournier E, Delmotte F, Dutech C. Population genetics of fungal diseases of plants. *Parasite* 2008;**15**:449−54.
13. Galagan JE, Henn MR, Ma LJ, Cuomo CA, Birren B. Genomics of the fungal kingdom: insights into eukaryotic biology. *Genome Res* 2005;**15**(12):1620−31.
14. Aguileta G, Hood M, Refrégier G, Giraud T. Genome evolution in pathogenic and symbiotic fungi. *Adv Bot Res* 2009;**49**:151−93.
15. Duplessis S, Cuomo CA, Lin Y-C, Aerts A, Tisserant E, Veneault-Fourrey C, et al. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc Natl Acad Sci USA* 2011;**108**(22):9166−71.
16. Kaemper J, Kahmann R, Boelker M, Ma L-J, Brefort T, Saville BJ, et al. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* 2006;**444**(7115): 97−101.
17. Schirawski J, Mannhaupt G, Muench K, Brefort T, Schipper K, Doehlemann G, et al. Pathogenicity determinants in smut fungi revealed by genome comparison. *Science* 2010; **330**(6010):1546−8.
18. Desjardins CA, Champion MD, Holder JW, Muszewska A, Goldberg J, Bailao AM, et al. Comparative genomic analysis of human fungal pathogens causing Paracoccidioidomycosis. *PLoS Genet* 2011;**7**(10).
19. Goodwin SB, Ben M'Barek S, Dhillon B, Wittenberg AHJ, Crane CF, Hane JK, et al. Finished genome of the fungal wheat pathogen *Mycosphaerella graminicola* reveals dispensome structure, chromosome plasticity, and stealth pathogenesis. *PLoS Genet* 2011;**7**(6).
20. Joneson S, Stajich JE, Shiu S-H, Rosenblum EB. Genomic transition to pathogenicity in chytrid fungi. *PLoS Pathog* 2011;**7**(11).
21. Pappas PG, Kauffman CA, Andes D, Benjamin Jr DK, Calandra TF, Edwards Jr JE, et al. Clinical practice guidelines for the management of candidiasis: 2009 update by the Infectious Diseases Society of America. *Clin Infect Dis* 2009;**48**(5):503−35.
22. Walsh TJ, Anaissie EJ, Denning DW, Herbrecht R, Kontoyiannis DP, Marr KA, et al. Treatment of aspergillosis: clinical practice guidelines of the Infectious Diseases Society of America. *Clin Infect Dis* 2008;**46**(3):327−60.
23. Marr KA, Carter RA, Boeckh M, Martin P, Corey L. Invasive aspergillosis in allogeneic stem cell transplant recipients: changes in epidemiology and risk factors. *Blood* 2002; **100**(13):4358−66.
24. Fraser JA, Stajich JE, Tarcha EJ, Cole GT, Inglis DO, Sil A, et al. Evolution of the mating type locus: insights gained from the dimorphic primary fungal pathogens *Histoplasma*

Downloaded From Vet-eBooks.com

capsulatum, Coccidioides immitis, and *Coccidioides posadasii*. *Eukaryot Cell* 2007;**6**(4): 622–9.
25. Sharpton TJ, Stajich JE, Rounsley SD, Gardner MJ, Wortman JR, Jordar VS, et al. Comparative genomic analyses of the human fungal pathogens *Coccidioides* and their relatives. *Genome Res* 2009;**19**(10):1722–31.
26. Reis RS, Almeida-Paes R, Muniz Mde M, Tavares PM, Monteiro PC, Schubach TM, et al. Molecular characterisation of *Sporothrix schenckii i*solates from humans and cats involved in the sporotrichosis epidemic in Rio de Janeiro, Brazil. *Mem Inst Oswaldo Cruz* 2009; **104**(5):769–74.
27. Hagen F, Khayhan K, Theelen B, Kolecka A, Polacheck I, Sionov E, et al. Recognition of seven species in the C*ryptococcus gattii/Cryptococcus neoformans* species complex. *Fungal Genet Biol* 2015;**78**:16–48.
28. Carlile M, Watkinson S, Gooday G. *The fungi*. 2nd ed. London: Academic Press; 2001. p. 588.
29. Sorrell TC. *Cryptococcus neoformans* variety *gattii*. *Med Mycol* 2001;**39**(2):155–68.
30. Kwon-Chung KJ, Boekhout T, Fell JW, Diaz M. Proposal to conserve the name *Cryptococcus gattii* against *C. hondurianus* and *C. bacillisporus* (Basidiomycota, Hymenomycetes, Tremellomycetidae). *Taxon* 2002;**51**(4):804–6.
31. Byrnes 3rd EJ, Bildfell RJ, Frank SA, Mitchell TG, Marr KA, Heitman J. Molecular evidence that the range of the Vancouver Island outbreak of *Cryptococcus gattii* infection has expanded into the Pacific Northwest in the United States. *J Infect Dis* 2009;**199**(7): 1081–6.
32. Fraser JA, Giles SS, Wenink EC, Geunes-Boyer SG, Wright JR, Diezmann S, et al. Same-sex mating and the origin of the Vancouver Island *Cryptococcus gattii* outbreak. *Nature* 2005;**437**(7063):1360–4.
33. Singh N, Alexander BD, Lortholary O, Dromer F, Gupta KL, John GT, et al. Pulmonary cryptococcosis in solid organ transplant recipients: clinical relevance of serum cryptococcal antigen. *Clin Infect Dis* 2008;**46**(2):e12–8.
34. MacDougall L, Kidd SE, Galanis E, Mak S, Leslie MJ, Cieslak PR, et al. Spread of *Cryptococcus gattii* in British Columbia, Canada, and detection in the Pacific Northwest, USA. *Emerg Infect Dis* 2007;**13**(1):42–50.
35. Lizarazo J, Linares M, de Bedout C, Restrepo A, Agudelo CI, Castaneda E. Results of nine years of the clinical and epidemiological survey on cryptococcosis in Colombia, 1997–2005. *Biomedica* 2007;**27**(1):94–109.
36. Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, et al. Emerging fungal threats to animal, plant and ecosystem health. *Nature* 2012; **484**(7393):186–94.
37. Olson DH, Aanensen DM, Ronnenberg KL, Powell CI, Walker SF, Bielby J, et al. Mapping the global emergence of *Batrachochytrium dendrobatidis*, the Amphibian chytrid fungus. *PLoS One* 2013;**8**(2).
38. Blehert D, Hicks A, Behr M, Meteyer C, Berlowski-Zier B, Buckles E, et al. Bat white-nose syndrome: an emerging fungal pathogen? *Science* 2009;**323**:227.
39. Meteyer C, Buckles E, Blehert D, Hicks A, Green D, Shearn-Bochsler V, et al. Histopathologic criteria to confirm white-nose syndrome in bats. *J Vet Diagn Invest* 2009;**21**: 411–4.
40. Byrnes 3rd EJ, Heitman J. *Cryptococcus gattii* outbreak expands into the Northwestern United States with fatal consequences. *F1000 Biol Rep* 2009;**1**(62):62.
41. Galanis E, MacDougall L. Epidemiology of *Cryptococcus gattii*, British Columbia, Canada, 1999–2007. *Emerg Infect Dis* 2010;**16**(2):251–7.

Downloaded From Vet-eBooks.com

42. Upton A, Fraser JA, Kidd SE, Bretz C, Bartlett KH, Heitman J, et al. First contemporary case of human infection with *Cryptococcus gattii* in Puget Sound: evidence for spread of the Vancouver Island outbreak. *J Clin Microbiol* 2007;**45**(9):3086−8.
43. Cheng PY, Sham A, Kronstad JW. *Cryptococcus gattii* isolates from the British Columbia cryptococcosis outbreak induce less protective inflammation in a murine model of infection than *Cryptococcus neoformans*. *Infect Immun* 2009;**77**(10):4284−94.
44. Abdolrasouli A, Rhodes J, Beale MA, Hagen F, Rogers TR, Chowdhary A, et al. Genomic context of azole resistance mutations in *Aspergillus fumigatus* determined using whole-genome sequencing. *Mbio* 2015;**6**(3).
45. Billmyre RB, Croll D, Li W, Mieczkowski P, Carter DA, Cuomo CA, et al. Highly recombinant VGII *Cryptococcus gattii* population develops clonal outbreak clusters through both sexual macroevolution and asexual microevolution. *Mbio* 2014;**5**(4).
46. Engelthaler DM, Hicks ND, Gillece JD, Roe CC, Schupp JM, Driebe EM, et al. *Cryptococcus gattii* in North American Pacific Northwest: whole-population genome analysis provides insights into species evolution and dispersal. *Mbio* 2014;**5**(4).
47. Farret RA, Desjardins CA, Sakthikumar S, Gujja S, Saif S, Zeng Q, et al. Genome evolution and innovation across the four major lineages of *Cryptococcus gattii*. *Mbio* 2015;**6**(5).
48. Datta K, Bartlett KH, Baer R, Byrnes E, Galanis E, Heitman J, et al. Spread of *Cryptococcus gattii* into Pacific Northwest region of the United States. *Emerg Infect Dis* 2009; **15**(8):1185−91.
49. Berger L, Speare R, Daszak P, Green D, Cunningham A, et al. Chytridiomycosis causes amphibian mortality associated with population declines in the rain forests of Australia and Central America. *Proc Natl Acad Sci USA* 1998;**95**:9031−6.
50. Longcore J, Pessier A, Nichols D. *Batrachochytrium dendrobatidis* gen. et sp. nov., a chytrid pathogenic to amphibians. *Mycologia* 1999;**91**:219−27.
51. Gleason F, Kagami M, Lefevre E, Sime-Ngando T. The ecology of chytrids in aquatic ecosystems: roles in food web dynamics. *Fungal Biol Rev* 2008:17−25.
52. Piotrowski J, Annis S, Longcore J. Physiology of *Batrachochytrium dendrobatidis*, a chytrid pathogen of amphibians. *Mycologia* 2004;**96**:9−15.
53. Fisher MC, Garner TWJ, Walker SF. Global emergence of *Batrachochytrium dendrobatidis* and amphibian chytridiomycosis in space, time, and host. *Annu Rev Microbiol* 2009;**63**:291−310.
54. Lips K, Diffendorfer J, Mendelson IJ, Sears M. Riding the wave: reconciling the roles of disease and climate change in amphibian declines. *PLoS Biol* 2008;**6**:441−54.
55. Lips K, Brem F, Brenes R, Reeve J, Alford R, et al. Emerging infectious disease and the loss of biodiversity in a Neotropical amphibian community. *Proc Natl Acad Sci USA* 2006; **103**:3165−70.
56. Weldon C, du Preez L, Hyatt A, Muller R, Speare R. Origin of the amphibian chytrid fungus. *Emerg Infect Dis* 2004;**10**:2100−5.
57. Goka K, Yokoyama J, Une Y, et al. Amphibian chytridiomycosis in Japan: distribution, haplotypes and possible route of entry into Japan. *Mol Ecol* 2009;**18**:4757−74.
58. Odds F, Bougnoux M, Shaw D, Bain J, Davidson A, et al. Molecular phylogenetics of *Candida albicans*. *Eukaryot Cell* 2007;**6**:1041−52.
59. Fisher M, Bosch J, Yin Z, Stead D, Walker J, et al. Proteomic and phenotypic profiling of an emerging pathogen of amphibians, *Batrachochytrium dendrobatidis*, shows that genotype is linked to virulence. *Mol Ecol* 2009;**18**:415−29.
60. Sampaio JP, Inacio J, Fonseca A, Gadanho M, Spencer-Martins I, Scorzetti G, et al. *Auriculibuller fuscus* gen. nov., sp. nov. and *Bullera japonica* sp. nov., novel taxa in the Tremellales. *Int J Syst Evol Microbiol* 2004;**54**(Pt 3):987−93.

Downloaded From Vet-eBooks.com

61. Ergin C, Ilkit M, Kaftanoglu O. Detection of *Cryptococcus neoformans* var. *grubii* in honeybee (*Apis mellifera*) colonies. *Mycoses* 2004;**47**(9−10):431−4.
62. Rimek D, Haase G, Luck A, Casper J, Podbielski A. First report of a case of meningitis caused by *Cryptococcus adeliensis* in a patient with acute myeloid leukemia. *J Clin Microbiol* 2004;**42**(1):481−3.
63. Findley K, Rodriguez-Carres M, Metin B, Kroiss J, Fonseca A, Vilgalys R, et al. Phylogeny and phenotypic characterization of pathogenic *Cryptococcus* species and closely related saprobic taxa in the Tremellales. *Eukaryot Cell* 2009;**8**(3):353−61.
64. Desprez-Loustau M, Robin C, Buee M, Courtecuisse R, Garbaye J, Suffert F, et al. The fungal dimension of biological invasions. *Trends Ecol Evol* 2007;**22**(9): 472−80.
65. Birch PRJ, Whisson SC. *Phytophthora infestans* enters the genomics era. *Mol Plant Pathol* 2001;**2**:257−63.
66. Stukenbrock EH, McDonald BA. The origins of plant pathogens in agro-ecosystems. *Annu Rev Phytopathol* 2008;**46**(1):75−100.
67. Philibert A, Desprez-Loustau M-L, Fabre B, Frey P, Halkett F, Husson C, et al. Predicting invasion success of forest pathogenic fungi from species traits. *J Appl Ecol* 2011;**48**(6): 1381−90.
68. Pimentel D, McNair S, Janecka J, Wightman J, Simmonds C, O'Connell C, et al. Economic and environmental threats of alien plant, animal, and microbe invasions. *Agric Ecosyst Environ* 2001;**84**(1):1−20.
69. Tibayrenc M. Genetic epidemiology of parasitic protozoa and other infectious agents: the need for an integrated approach. *Int J Parasitol* 1998;**28**(1):85−104.
70. Taylor JW, Geiser DM, Burt A, Koufopanou V. The evolutionary biology and population genetics underlying fungal strain typing. *Clin Microbiol Rev* 1999;**12**(1): 126−46.
71. Tibayrenc M. Bridging the gap between molecular epidemiologists and evolutionists. *Trends Microbiol* 2005;**13**(12):575−80.
72. Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* 1998;**95**(6):3140−5.
73. Fisher MC, Koenig GL, White TJ, San-Blas G, Negroni R, Alvarez IG, et al. Biogeographic range expansion into South America by *Coccidioides immitis* mirrors New World patterns of human migration. *Proc Natl Acad Sci USA* 2001;**98**(8):4558−62.
74. James TY, Litvintseva AP, Vilgalys R, Morgan JAT, Taylor JW, Fisher MC, et al. Rapid global expansion of the fungal disease chytridiomycosis into declining and healthy amphibian populations. *PLoS Pathog* 2009;**5**(5).
75. Fisher MC, Koenig GL, White TJ, Taylor JW. Pathogenic clones versus environmentally driven population increase: analysis of an epidemic of the human fungal pathogen *Coccidioides immitis*. *J Clin Microbiol* 2000;**38**(2):807−13.
76. Ivors KL, Hayden KJ, Bonants PJM, Rizzo DM, Garbelotto M. AFLP and phylogenetic analyses of North American and European populations of *Phytophthora ramorum*. *Mycol Res* 2004;**108**:378−92.
77. Prospero S, Hansen EM, Grunwald NJ, Winton LM. Population dynamics of the sudden oak death pathogen *Phytophthora ramorum* in Oregon from 2001 to 2004. *Mol Ecol* 2007; **16**(14):2958−73.
78. Mascheretti S, Croucher PJP, Vettraino A, Prospero S, Garbelotto M. Reconstruction of the Sudden Oak Death epidemic in California through microsatellite analysis of the pathogen *Phytophthora ramorum*. *Mol Ecol* 2008;**17**(11):2755−68.

Downloaded From Vet-eBooks.com

79. Peever TL, Salimath SS, Su G, Kaiser WJ, Muehlbauer FJ. Historical and contemporary multilocus population structure of *Ascochyta rabiei* (teleomorph : *Didymella rabiei*) in the Pacific Northwest of the United States. *Mol Ecol* 2004;**13**(2):291−309.
80. McDonald BA, Linde C. Pathogen population genetics, evolutionary potential, and durable resistance. *Annu Rev Phytopathol* 2002;**40**:349−79.
81. Giraud T, Gladieux P, Gavrilets S. Linking emergence of fungal plant diseases and ecological speciation. *Trends Ecol Evol* 2010;**25**:387−95.
82. Williams PD. Darwinian interventions: taming pathogens through evolutionary ecology. *Trends Parasitol* 2009;**26**(2):83−92.
83. Taylor JW, Turner E, Townsend JP, Dettman JR, Jacobson D. Eukaryotic microbes, species recognition and the geographic limits of species: examples from the kingdom Fungi. *Philos Trans R Soc Lond B-Biol Sci* 2006;**361**(1475):1947−63.
84. Otto SP. The evolutionary enigma of sex. *Am Nat* 2009;**174**:S1−14.
85. Milgroom MG. Recombination and the multilocus structure of fungal populations. *Annu Rev Phytopathol* 1996;**34**(1):457−77.
86. Giraud T, Fortini D, Levis C, Leroux P, Brygoo Y. RFLP markers show genetic recombination in *Botryotinia fuckeliana* (*Botrytis cinerea*) and transposable elements reveal two sympatric species. *Mol Biol Evol* 1997;**14**(11):1177−85.
87. Maynard-Smith J, Smith NH, O'Rourke M, Spratt BG. How clonal are bacteria? *Proc Natl Acad Sci USA* 1993;**90**(10):4384−8.
88. Fisher MC, Hanage WP, de Hoog S, Johnson E, Smith MD, White NJ, et al. Low effective dispersal of asexual genotypes in heterogeneous landscapes by the endemic pathogen *Penicillium marneffei*. *PLoS Pathog* 2005;**1**(2):159−65.
89. Burt A, Carter DA, Koenig GL, White TJ, Taylor JW. Molecular markers reveal cryptic sex in the human pathogen *Coccidioides immitis*. *Proc Natl Acad Sci USA* 1996;**93**(2): 770−3.
90. Mboup M, Leconte M, Gautier A, Wan AM, Chen W, de Vallavieille-Pope C, et al. Evidence of genetic recombination in wheat yellow rust populations of a Chinese oversummering area. *Fungal Genet Biol* 2009;**46**(4):299−307.
91. Hudson RR. Estimating the recombination parameter of a finite population-model without selection. *Genet Res* 1987;**50**(3):245−50.
92. Broquet T, Petit EJ. Molecular estimation of dispersal for ecology and population genetics. *Annu Rev Ecol Evol Syst* 2009;**40**:193−216.
93. Slatkin M. Gene-flow in natural populations. *Annu Rev Ecol Syst* 1985;**16**:393−430.
94. Yarwood CE. Man-made plant diseases. *Science* 1970;**168**:218−20.
95. Bossart JL, Prowell DP. Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends Ecol Evol* 1998;**13**(5):202−6.
96. Whitlock MC, McCauley DE. Indirect measures of gene flow and migration: F-ST not equal 1/(4Nm + 1). *Heredity* 1999;**82**:117−25.
97. Gladieux P, Zhang X-G, Afoufa-Bastien D, Valdebenito Sanhueza R-M, Sbaghi M, Le Cam B. On the origin and spread of the scab disease of apple: out of central Asia. *PLoS One* 2008;**3**(1):e1455.
98. Gladieux P, Zhang XG, Roldan-Ruiz I, Caffier V, Leroy T, Devaux M, et al. Evolution of the population structure of *Venturia inaequalis*, the apple scab fungus, associated with the domestication of its host. *Mol Ecol* 2010;**19**(4):658−74.
99. Barres B, Halkett F, Dutech C, Andrieux A, Pinon J, Frey P. Genetic structure of the poplar rust fungus *Melampsora larici-populina*: evidence for isolation by distance in Europe and recent founder effects overseas. *Infect Genet Evol* 2008;**8**(5):577−87.
100. Kingman JFC. The coalescent. *Stoch Process Appl* 1982;**13**:235−48.

Downloaded From Vet-eBooks.com

101. Kuhner MK. Coalescent genealogy samplers: windows into population history. *Trends Ecol Evol* 2009;**24**(2):86−93.
102. Stephens M. Inference under the coalescent. In: Balding DJ, Bishop M, Cannings C, editors. *Handbook of statistical genetics*. 3rd ed. Chichester: John Wiley & Sons, Ltd.; 2008. p. 878−908.
103. Hey J, Nielsen R. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 2004;**167**(2):747−60.
104. Hey J, Nielsen R. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc Natl Acad Sci USA* 2007;**104**(8): 2785−90.
105. Zaffarano PL, McDonald BA, Linde CC. Phylogeographical analyses reveal global migration patterns of the barley scald pathogen *Rhynchosporium secalis*. *Mol Ecol* 2009; **18**(2):279−93.
106. Beaumont MA, Zhang WY, Balding DJ. Approximate Bayesian computation in population genetics. *Genetics* 2002;**162**(4):2025−35.
107. Guillemaud T, Beaumont MA, Ciosi M, Cornuet JM, Estoup A. Inferring introduction routes of invasive species using approximate Bayesian computation on microsatellite data. *Heredity* 2010;**104**(1):88−99.
108. Miller N, Estoup A, Toepfer S, Bourguet D, Lapchin L, Derridj S, et al. Multiple transatlantic introductions of the western corn rootworm. *Science* 2005;**310**(5750): 992.
109. Cornuet JM, Santos F, Beaumont MA, Robert CP, Marin JM, Balding DJ, et al. Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* 2008;**24**(23):2713−9.
110. Rousset F. Genetic differentiation between individuals. *J Evol Biol* 2000;**13**(1):58−62.
111. Dutech C, Rossi JP, Fabreguettes O, Robin C. Geostatistical genetic analysis for inferring the dispersal pattern of a partially clonal species: example of the chestnut blight fungus. *Mol Ecol* 2008;**17**(21):4597−607.
112. Nielsen R, Wakeley J. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* 2001;**158**(2):885−96.
113. Stukenbrock EH, Banke S, Javan-Nikkhah M, McDonald BA. Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation. *Mol Biol Evol* 2007;**24**(2):398−411.
114. Staats M, van Baarlen P, van Kan JAL. Molecular phylogeny of the plant pathogenic genus *Botrytis* and the evolution of host specificity. *Mol Biol Evol* 2005;**22**(2):333−46.
115. Gladieux P, Caffier V, Devaux M, Le Cam B. Host-specific differentiation among populations of *Venturia inaequalis* causing scab on apple, pyracantha and loquat. *Fungal Genet Biol* 2010;**47**:511−21.
116. Kasuga T, White TJ, Koenig G, McEwen J, Restrepo A, Castaneda E, et al. Phylogeography of the fungal pathogen *Histoplasma capsulatum*. *Mol Ecol* 2003;**12**(12):3383−401.
117. Bleykasten-Grosshans C, Neuveglise C. Transposable elements in yeasts. *C R Biol* 2011; **334**(8−9):679−86.
118. Brefort T, Tanaka S, Neidig N, Doehlemann G, Vincon V, Kahmann R. Characterization of the largest effector gene cluster of *Ustilago maydis*. *PLoS Pathog* 2014;**10**(7).
119. Holder M, Lewis PO. Phylogeny estimation: traditional and Bayesian approaches. *Nat Rev Genet* 2003;**4**(4):275−84.
120. Hey J, Machado CA. The study of structured populations − new hope for a difficult and divided science. *Nat Rev Genet* 2003;**4**(7):535−43.

Downloaded From Vet-eBooks.com

121. Posada D, Crandall KA. Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol Evol* 2001;**16**(1):37−45.
122. Guillot G, Leblois R, Coulon A, Frantz AC. Statistical methods in spatial genetics. *Mol Ecol* 2009;**18**(23):4734−56.
123. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 2003;**164**(4): 1567−87.
124. Jombart T, Pontier D, Dufour AB. Genetic markers in the playground of multivariate analysis. *Heredity* 2009;**102**(4):330−41.
125. Cavalli-Sforza LL. Population structure and human evolution. *Proc R Soc Lond Ser B Biol Sci* 1966;**164**(995):362−79.
126. Dutech C, Fabreguettes O, Capdevielle X, Robin C. Multiple introductions of divergent genetic lineages in an invasive fungal pathogen, *Cryphonectria parasitica*, in France. *Heredity* 2009.
127. Heitman J, Kronstad JW, Taylor JW, Casselton LA. *Sex in fungi: molecular determination and evolutionary implications*. Washington DC: American Society for Microbiology Press; 2007.
128. Billiard S, López-Villavicencio M, Devier B, Hood ME, Fairhead C, Giraud T. Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. *Biol Rev* 2011;**86**:421−42.
129. Giraud T, Refregier G, Le Gac M, de Vienne DM, Hood ME. Speciation in fungi. *Fungal Genet Biol* 2008;**45**(6):791−802.
130. Kohn LM. Mechanisms of fungal speciation. *Annu Rev Phytopathol* 2005;**43**(1):279−308.
131. Stukenbrock EH, Croll D. The evolving fungal genome. *Fungal Biol Rev* 2014;**28**(1): 1−12.
132. Gladieux P, Ropars J, Badouin H, Branca A, Aguileta G, De Vienne DM, et al. Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. *Mol Ecol* 2014;**23**(4):753−73.
133. Dean R, Van Kan JAL, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, Spanu PD, et al. The top 10 fungal pathogens in molecular plant pathology. *Mol Plant Pathol* 2012;**13**(4): 414−30.
134. Richardson MD. Opportunistic and pathogenic fungi. *J Antimicrob Chemother* 1991;**28**: 1−11.
135. Pfaller MA, Diekema DJ. Rare and emerging opportunistic fungal pathogens: concern for resistance beyond *Candida albicans* and *Aspergillus fumigatus*. *J Clin Microbiol* 2004; **42**(10):4419−31.
136. Soanes DM, Alam I, Cornell M, Wong HM, Hedeler C, Paton NW, et al. Comparative genome analysis of filamentous fungi reveals gene family expansions associated with fungal pathogenesis. *PLoS One* 2008;**3**(6).
137. Moran GP, Coleman DC, Sullivan DJ. Comparative genomics and the evolution of pathogenicity in human pathogenic fungi. *Eukaryot Cell* 2011;**10**(1):34−42.
138. Morel G, Sterck L, Swennen D, Marcet-Houben M, Onesime D, Levasseur A, et al. Differential gene retention as an evolutionary mechanism to generate biodiversity and adaptation in yeasts. *Sci Rep* 2015;**5**:11571.
139. Cuomo CA, Gueldener U, Xu JR, Trail F, Turgeon BG, Di Pietro A, et al. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* 2007;**317**(5843):1400−2.
140. Xu J, Saunders CW, Hu P, Grant RA, Boekhout T, Kuramae EE, et al. Dandruff-associated *Malassezia* genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens. *Proc Natl Acad Sci USA* 2007;**104**(47):18730−5.

Downloaded From Vet-eBooks.com

141. Pendleton AL, Smith KE, Feau N, Martin FM, Grigoriev IV, Hamelin R, et al. Duplications and losses in gene families of rust pathogens highlight putative effectors. *Front Plant Sci* 2014:5.
142. Soanes DM, Richard T, Talbot N. Insights from sequencing fungal and oomycete genomes: what can we learn about plant disease and the evolution of pathogenicity? *Plant Cell* 2007;**19**:3318−26.
143. Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, et al. Genome evolution in yeasts. *Nature* 2004;**430**(6995):35−44.
144. Kellis M, Birren BW, Lander ES. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 2004;**428**(6983):617−24.
145. Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* 2006;**440**(7082):341−5.
146. Verstrepen KJ, Fink GR. Genetic and epigenetic mechanisms underlying cell-surface variability in protozoa and fungi. *Annu Rev Genet* 2009;**43**:1−24.
147. Gabaldon T, Martin T, Marcet-Houben M, Durrens P, Bolotin-Fukuhara M, Lespinet O, et al. Comparative genomics of emerging pathogens in the Candida glabrata clade. *BMC Genomics* 2013:14.
148. Howlett BJ. Secondary metabolite toxins and nutrition of plant pathogenic fungi. *Curr Opin Plant Biol* 2006;**9**(4):371−5.
149. Gbelska Y, Krijger JJ, Breunig KD. Evolution of gene families: the multidrug resistance transporter genes in five related yeast species. *FEMS Yeast Res* 2006;**6**(3):345−55.
150. Chen W, Lee M-K, Jefcoate C, Kim S-C, Chen F, Yu J-H. Fungal cytochrome P450 monooxygenases: their distribution, structure, functions, family expansion, and evolutionary origin. *Genome Biol Evol* 2014;**6**(7):1620−34.
151. Marcet-Houben M, Marceddu G, Gabaldon T. Phylogenomics of the oxidative phosphorylation in fungi reveals extensive gene duplication followed by functional divergence. *BMC Evol Biol* 2009;**9**:295.
152. Rech GE, Sanz-Martin JM, Anisimova M, Sukno SA, Thon MR. Natural selection on coding and noncoding DNA sequences is associated with virulence genes in a plant pathogenic fungus. *Genome Biol Evol* 2014;**6**(9):2368−79.
153. Staats M, van Baarlen P, Schouten A, van Kan JAL, Bakker FT. Positive selection in phytotoxic protein-encoding genes of *Botrytis* species. *Fungal Genet Biol* 2007;**44**(1): 52−63.
154. Meyers BC, Chin DB, Shen KA, Sivaramakrishnan S, Lavelle DO, Zhang Z, et al. The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. *Plant Cell* 1998;**10**(11):1817−32.
155. Jargeat P, Rekangalt D, Verner MC, Gay G, Debaud JC, Marmeisse R, et al. Characterisation and expression analysis of a nitrate transporter and nitrite reductase genes, two members of a gene cluster for nitrate assimilation from the symbiotic basidiomycete *Hebeloma cylindrosporum*. *Curr Genet* 2003;**43**(3):199−205.
156. Mattern DJ, Schoeler H, Weber J, Novohradska S, Kraibooj K, Dahse H-M, et al. Identification of the antiphagocytic trypacidin gene cluster in the human-pathogenic fungus *Aspergillus fumigatus*. *Appl Microbiol Biotechnol* 2015;**99**(23):10151−61.
157. Lawler K, Hammond-Kosack K, Brazma A, Coulson RMR. Genomic clustering and coregulation of transcriptional networks in the pathogenic fungus *Fusarium graminearum*. *BMC Syst Biol* 2013:7.
158. Rehmeyer C, Li WX, Kusaba M, Kim YS, Brown D, Staben C, et al. Organization of chromosome ends in the rice blast fungus, *Magnaporthe oryzae*. *Nucleic Acids Res* 2006; **34**(17):4685−701.

Downloaded From Vet-eBooks.com

159. Kang S, Lebrun MH, Farrall L, Valent B. Gain of virulence caused by insertion of a Pot3 transposon in a *Magnaporthe grisea* avirulence gene. *Mol Plant-Microbe Interact* 2001; **14**(5):671−4.
160. Fudal I, Ross S, Gout L, Blaise F, Kuhn ML, Eckert MR, et al. Heterochromatin-like regions as ecological niches for avirulence genes in the *Leptosphaeria maculans* genome: map-based cloning of AvrLm6. *Mol Plant-Microbe Interact* 2007;**20**(4):459−70.
161. Gout L, Fudal I, Kuhn ML, Blaise F, Eckert M, Cattolico L, et al. Lost in the middle of nowhere: the AvrLm1 avirulence gene of the Dothideomycete *Leptosphaeria maculans*. *Mol Microbiol* 2006;**60**(1):67−80.
162. Sánchez-Alonso P, Guzman P. Predicted elements of telomere organization and function in *Ustilago maydis*. *Fungal Genet Biol* 2008;**45**(Suppl. 1):S54−62.
163. Chen QH, Wang YC, Li AN, Zhang ZG, Zheng XB. Molecular mapping of two cultivar-specific avirulence genes in the rice blast fungus *Magnaporthe grisea*. *Mol Genet Genomics* 2007;**277**(2):139−48.
164. Penalva MA, Moya A, Dopazo J, Ramon D. Sequences of isopenicillin-N synthetase genes suggest horizontal gene-transfer from prokaryotes to eukaryotes. *Proc R Soc Lond Ser B-Biol Sci* 1990;**241**(1302):164−9.
165. Soanes D, Richards TA. Horizontal gene transfer in eukaryotic plant pathogens. In: VanAlfen NK, editor. *Annual review of phytopathology*, vol. 52; 2014. p. 583−614. 522014.
166. Garcia-Vallve S, Romeu A, Palau J. Horizontal gene transfer of glycosyl hydrolases of the rumen fungi. *Mol Biol Evol* 2000;**17**(3):352−61.
167. Gojkovic Z, Knecht W, Zameitat E, Warneboldt J, Coutelis JB, Pynyaha Y, et al. Horizontal gene transfer promoted evolution of the ability to propagate under anaerobic conditions in yeasts. *Mol Genet Genomics* 2004;**271**(4):387−93.
168. Cheeseman K, Ropars J, Renault P, Dupont J, Gouzy J, Branca A, et al. Multiple recent horizontal transfers of a large genomic region in cheese making fungi. *Nat Commun* 2014:5.
169. Ropars J, Rodriguez de la Vega RC, Lopez-Villavicencio M, Gouzy J, Sallet E, Dumas E, et al. Adaptive horizontal gene transfers between multiple cheese-associated fungi. *Curr Biol* 2015;**25**(19):2562−9.
170. Friesen TL, Stukenbrock EH, Liu Z, Meinhardt S, Ling H, Faris JD, et al. Emergence of a new disease as a result of interspecific virulence gene transfer. *Nat Genet* 2006;**38**(8): 953−6.
171. Walton JD. Horizontal gene transfer and the evolution of secondary metabolite gene clusters in fungi: an hypothesis. *Fungal Genet Biol* 2000;**30**(3):167−71.
172. Pryszcz LP, Nemeth T, Gacser A, Gabaldon T. Genome comparison of *Candida orthopsilosis* clinical strains reveals the existence of hybrids between two distinct subspecies. *Genome Biol Evol* 2014;**6**(5):1069−78.
173. Casadevall A. Evolution of intracellular pathogens. *Annu Rev Microbiol* 2008;**62**:19−33.
174. Morris CE, Bardin M, Kinkel LL, Moury B, Nicot PC, Sands DC. Expanding the paradigms of plant pathogen life history and evolution of parasitic fitness beyond agricultural boundaries. *PLoS Pathog* 2009;**5**(12):e1000693.
175. Desprez-Loustau ML, Courtecuisse R, Robin C, Husson C, Moreau PA, Blancard D, et al. Species diversity and drivers of spread of alien fungi (sensu lato) in Europe with a particular focus on France. *Biol Invasions* 2010;**12**(1):157−72.
176. Sexton AC, Howlett BJ. Parallels in fungal pathogenesis on plant and animal hosts. *Eukaryot Cell* 2006;**5**(12):1941−9.
177. Dickman MB, de Figueiredo P. Comparative pathobiology of fungal pathogens of plants and animals. *PLoS Pathog* 2011;**7**(12).

Downloaded From Vet-eBooks.com

178. Sharon A, Shlezinger N. Fungi infecting plants and animals: killers, non-killers, and cell death. *PLoS Pathog* 2013;**9**(8).
179. Ottmann C, Luberacki B, Kufner I, Koch W, Brunner F, Weyand M, et al. A common toxin fold mediates microbial attack and plant defense. *Proc Natl Acad Sci USA* 2009; **106**(25):10359−64.
180. Jarne P, Lagoda PJL. Microsatellites, from molecules to populations and back. *Trends Ecol Evol* 1996;**11**:424−9.
181. Dutech C, Enjalbert E, Fournier E, Delmotte F, Barrès B, Carlier J, et al. Challenges of microsatellite isolation in fungi. *Fungal Genet Biol* 2007;**44**:933−49.
182. Balloux F, Lugon-Moulin N. The estimation of population differentiation with microsatellite markers. *Mol Ecol* 2002;**11**:155−65.
183. Halkett F, Simon J-C, Balloux F. Tackling the population genetics of clonal and partially clonal organisms. *Trends Ecol Evol* 2005;**20**:194−201.

Downloaded From Vet-eBooks.com

Clonal Evolution

5

T. de Meeûs[1], *F. Prugnolle*[2]
[1]UMR 177 IRD — CIRAD INTERTRYP, Campus International de Baillarguet, Montpellier, France; [2]Maladies Infectieuses et Vecteurs Ecologie, Génétique, Evolution et Contrôle, MIVEGEC (IRD 224-CNRS 5290-UM1-UM2), IRD Center, Montpellier, France

1. Introduction

Asexual reproduction is probably the most widespread means of biological propagation[1,2] and is probably the oldest one though recombination might be almost as old.[3] But this of course depends on what is meant and what is understood (not always the same thing) by clonality and recombination.

Asexual reproduction has been the subject of numerous studies and reviews from diverse biological disciplines.[1,2,4–9] The issue appears to be perceived differently for specialists working on Bacteria, Archaea, Eukaryota, unicellular or pluricellular animals, or plants. In this chapter, we therefore first deal with specific definitions as this subject area is littered with vocabulary that sometimes has ambiguous meanings. We then try to go back in time to the origin of asexual reproduction and recombination and attempt to describe the diversity of ways in which prokaryotes and eukaryotes reproduce asexually and recombine. Following this we describe the various ways that asexual reproduction is incorporated in eukaryotic life cycles. After a brief attempt to quantify the importance of asexuality in living organisms, the genetic consequences of asexuality are reviewed, followed by a section on the evolution and the paradox of sex. What evolutionary advantages are brought by clonality? What disadvantages result from clonality? What is the so-called twofold cost of sex? The last section deals with clonal microevolution. It consists of two parts: the first one treats neutral gene variability in clonal populations (population genetics structure) and the second addresses selective issues, such as the evolution of resistance or virulence in clonal populations. Finally, we conclude with economic and medical issues linked to asexual organisms.

2. Definitions

Asexual reproduction is a process of genetic propagation of genomes, following which the genomes that descend from this process are strictly identical to the parental genome, in terms of quantity and quality, at the exception of uncorrected errors during the duplication process (i.e., mutations).[1] Besides cell division (e.g., mitosis in unicellular eukaryotes), many other processes correspond to clonal propagation as

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00005-6
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

agametic (animals) or vegetative (plants) reproduction, ameiotic thelytokous parthenogenesis, endomitotic automictic parthenogenesis with pair formation of sister chromatids occurring before meiosis, automictic parthenogenesis with fusion of two polar bodies, deuterokous parthenogenesis, gynogenesis, apomixy, or agamospermy (reviewed in Ref. 1).

Sexual reproduction is not initially a propagation mode even if it is now 100% correlated with the multiplication of many organisms (e.g., mammals). It is a recombinational repair tool,[3,10,11] hence the use of sexual recombination (SR) in the rest of this chapter as a synonym for meiotic sex. Recombination in the wide sense is present in the three domains of life (Archaea, Bacteria, and Eukaryota), although through very different means,[3] while SR is a eukaryotic hallmark.[3,10,12] Recombination can take three forms in Bacteria and Archaea: conjugation, transformation, and transduction.[3,13,14] Conjugation concerns plasmid exchange through a specialized structure called pilus. It is unidirectional in Bacteria (donor and recipient) and is apparently bidirectional in certain Euryarchaeota.[14] Transformation is the absorption of soluble naked DNA present in the microenvironment by a recipient cell, and its further inclusion (recombination), if compatible, in the chromosome. When divergence between the two sequences is less than 25%, a homologous recombination can occur (without chromosome size increase). Illegitimate integration of more divergent DNA can increase the size of the recipient chromosome. Homology-based recombination also increases recipient chromosome size. Natural transformation can be found in any eubacteria lineage, but has only been reported in 1% of recognized eubacteria species (see Ref. 15 for review) and was never observed in Archaea except in laboratory conditions, by chemical or physical induction.[14] Transduction is a horizontal gene transfer mediated by viruses. It is widespread in Eubacteria,[16] while in Archaea, it has been reported in methanogens (Euryarchaeota) only.[14] Calling transduction, transformation, and conjugation sex is unsound and true sex, with meiosis and syngamy, is only found in eukaryotes and never in prokaryotes.[3]

Panmixia defines a population where zygotes (eggs) are produced by the random syngamy (union) of available sexual cells. It can thus only occur in eukaryotes if any. Then, talking about panmictic bacteria is inappropriate as well. The genetic consequence of panmixia is the establishment of the famous Hardy−Weinberg (HW) genotypic proportions of the form p^2, $2pq$, and q^2 (for two alleles of frequencies p and q). These proportions are only expected to be approximately met in populations of highly mobile monoecious individuals with panmictic sex. Consequently, talking of panmixia for a microbe is also fairly unsound.

Linkage disequilibrium (LD) reflects the statistical association between different alleles at different loci in the genome. LD can be generated by virtually all evolutionary forces. Besides the obvious physical linkage, selection, population structure (small subpopulation sizes and migration), mutation, and reproductive system (except panmixia) all have a positive impact on LD. Estimation and testing of positive LD is a hard task, and only very strong signals are expected to be detected, the variance of which is expected to be substantial.[17,18] Furthermore, very strong interactions between sampling design, reproductive system, and population structure can

Downloaded From Vet-eBooks.com

considerably bias LD perception.[19] Consequently, assessing reproductive systems through LD measures is at best risky, and measuring it through the proportion of significant LD tests found is definitely flawed. Small panmictic populations are expected to display high levels of HD.

3. The Origin of Life, the Origin of Propagation, and Recombination

Whether an RNA phase came before the DNA world is not discussed here, though now much evidence advocate for a life on earth that arose from an RNA World.[20] There is also a large consensus on the fact that all extant life is the descent of a single ancestor.[21] The last universal common ancestor (LUCA), also known as the cenancestor,[3] originated some 3−3.5 billion years ago.[22] The emergence of LUCA has probably followed a phase of extensive horizontal gene transfer (HGT) between the different arising entities.[10,21] The order of branching of Bacteria, Eukaryota, and Archaea domains is controversial, one interesting hypothesis being that eukaryotes emerged as the result of a symbiotic fusion of some bacterial and archaeal lineages.[23] Confusion finds its origin in the potentially important, disturbing HGT believed to, occasionally or often, occur between prokaryotic organisms.[24] Evolution of meiosis is viewed by certain as a defense mechanism that evolved against HGT to promote the best coordination between coevolved functions. When chromosomes pair during meiosis, a number of mechanisms such as repair, conversion, and recombination are triggered, allowing the elimination of deleterious differences, which is viewed as a protection against HGT.[10] Nevertheless, meiosis probably arose from mitosis, which is also specific to eukaryotes.[3] According to this author, SR appeared about 850 million years ago as a cell cycle repair mechanism to correct accidental polyploidy. Many of enzymes involved in meiosis have related enzymes in prokaryotic tool kits for controlling replication fidelity (rescue of broken or stalled replication forks, recombination, or mismatch corrections).[3,12]

Consequently, clonality evolved first (whether prokaryotes appeared first or not), but recombination probably arose soon after or at the same time to control for intensive HGT and/or polyploidy, and this was followed then by SR in eukaryotes. It is noteworthy that SR emergence is not presented as a response to a changing environment (red queen hypothesis) or to prevent Muller's ratchet of deleterious allele accumulation (e.g., see Refs. 1,8 for review) but as a mechanism for restoring genomic harmony after replication mistakes or any DNA damage. The fact SR did not evolve in prokaryotes probably comes from the constraints resulting from their particular peptidoglycan envelope said to act as a "chastity belt."[3] It is nevertheless a proof that SR is by no means a necessity to adapt to variable environments or fight against Muller's ratchet.

Microbes represent the major part of genetic diversity on earth, most of which is still represented by uncultivated organisms.[24] Clonality is thus as old as

Downloaded From Vet-eBooks.com

life and widespread on earth. It does not evolve in competition with recombination in the wide sense (it being sexual or not) but coevolves with it in most situations.

4. Clonal Modes

As seen, prokaryotes have various ways to recombine and only one way to divide.[3] On the contrary, eukaryotes, and in particular pluricellular ones, have barely a single way for recombination (if we exclude possible gene transfer through viruses or with endosymbionts) and many different ways to propagate clonally. Reviewing all these modes would be tedious and unnecessary as most was already presented in a 2007 review.[1] It is interesting though to focus briefly on a particular family of clonal modes that diverted SR to, so to speak, reintegrate back clonal reproduction. The different forms of parthenogenesis that produce daughters identical to their mother (see earlier section) correspond to that. These different parthenogenesis modes are obviously those that attracted most attention of evolutionary biologists working on the evolution of sex, in particular the famous asexual scandal of bdelloid rotifers.[25,26] In fact, fixed clonality has rarely been demonstrated, but the coexistence of both systems is much more the rule as in aphids, other rotifers (except purely sexual acanthocephalans), cycliophorans, and many others.[1] The fact that it must have been a real challenge to divert meiosis apparatus and that this nevertheless evolved many times in complex eukaryotes appears as a spectacular illustration of how costly SR must be, hence the impressive amount of works dedicated to this issue (see following sections).

De Meeûs et al.[1,2] found it convenient to classify organisms according to the kind of cycle these are involved in with regard to clonal propagation. We stick to this classification in the following. This classification separates four kinds of cycles: (1) the purely sexual cycle (Sex) corresponds to organisms that can only reproduce through SR; (2) complex life cycles with an instantaneous clonal phase with only one (I) clonal generation per cycle; (3) complex life cycles with several generations of asexuality (S) where the clonal phase involves more than one clonal generation, and finally (4) life cycles where sexual reproduction is more or less frequent (or even absent) with an acyclic pattern (A). In cases 2 and 3 (i.e., I and S), and for all surviving individuals, SR must intervene at one point in the cycle to form zygotes. In case 4 (A), the life cycle is not defined by a regular pattern of sexual or asexual reproduction. Case 1 (Sex) is typical of vertebrates, especially mammals and birds but also cestodes, most arthropods, or nematodes. Cycle 2 (I) applies to all species with polyembryony and many budding species. For example, this cycle is typical of trematodes (flukes). Case 3 (S) is typical of aphids, monogonont rotifers, cladocerans, many fungi, and most Sporozoa (parasitic unicellular organisms, including the malaria agents *Plasmodium* spp.), and probably *Leishmania* (see Ref. 27). Finally, case 4 (A) is common in plants and unicellular organisms. In particular, it is found in strictly clonal organisms, or at least those organisms in which sex is unknown, such as bdelloid rotifers, imperfect fungi (e.g., *Candida albicans*), Parabasalia (*Trichomonas vaginalis*),

Downloaded From Vet-eBooks.com

Metamonadina (*Giardia lamblia*), parasitic amoebas, and *Trypanozoma brucei gambiense*, the agent of sleeping sickness.

5. Quantifying the Importance of Asexuality in the Biosphere

There are two ways to comprehend this issue. In terms of described (known) species, purely sexual species are the most represented.[2] Nevertheless, there is an obvious bias in accounting biological diversity through described species.[28,29] As quoted earlier, microbes (cycles S or A) represent the major part of genetic diversity on earth, most of which is still represented by uncultivated organisms.[24] It can thus be safely postulated that organisms with a clonal phase represent the major part of biodiversity. If this was accounted for in terms of energy devoted to clonality and SR on earth per second, SR would probably look like an epiphenomenon. This should be trivial as the real way to propagate life is through cell (hence asexual) division, while SR is in fact meant for DNA repair and/or to control DNA replication fidelity.

The numeric importance of clonal parasitic eukaryotes was already reviewed.[2] Whole described species again give a biased advantage to purely sexual species. Nevertheless, a glance at the most documented human parasitic fauna completely reverses the tendency thus suggesting: (1) that parasite represent the most important part of eukaryotic biodiversity and (2) that clonal species (i.e., using this mode at one stage of their life cycle) are in majority among them. If Archaea and Bacteria are included, known species number is useless. There are indeed more known bird species than the sum of known Archaea and Bacteria, which is nonsense. Prokaryotes are so numerous everywhere that estimating how much of its diversity specialized in parasitism looks like an unreachable chimera. We can however suspect this number to be tremendous regarding all bacterial diseases that can affect mankind (around 43 after a quick and dirty look in the web). For eukaryotic parasites alone, it was estimated that more than a billion people are affected by such kind of diseases,[2] some of which figure as the most severe ones (e.g., malaria). Clonality in infectious disease cannot thus be treated lightly.

6. Genetic Consequences of Asexuality

This issue was reviewed many times (e.g., in Refs. 1,2,7,9,25,30–43), so we will be brief and stick to the essential. In haploid organisms, clonality tends to increase statistical associations between the different loci of the genome irrespective of their location. In pure asexuals, this should end up with the presence of numerous repetitions of certain clones, hence of the same multilocus genotypes (MLGs). Depending on population structure, MLG diversity will vary from low (e.g., a single MLG) to high variability (several MLGs). As linkage is total, MLGs can be considered as the different alleles of a single locus. If no SR is involved, it is expected that the different MLGs

Downloaded From Vet-eBooks.com

that can be maintained can potentially be highly divergent. This may represent a problem because at a given level of divergence, it is probable that adaptive differences will arise. Moreover, especially in small subpopulations that are not expected to maintain much equivalent different MLGs, the stable maintenance of highly diverged MLGs of the same "species" might reflect more an ecological divergence, for example, the coexistence of different ecotype or species, than a simple genetic polymorphism. When some SR is involved, the combination between drift, reproduction, and sampling renders difficult the interpretation of the patterns of genetic variability in haploids. This is also true for diploids even if heterozygosity can be helpful to that respect. When the amount of SR is large enough, populations display patterns of genetic variability close to that observed for purely sexual (but not necessarily panmictic) populations.

In diploids, haplotypic consequences are similar but here, in the absence of SR, the two alleles of a lineage will continuously diverge since the last SR event. Consequently, as illustrated in Fig. 5.1, divergence between the two alleles of the same

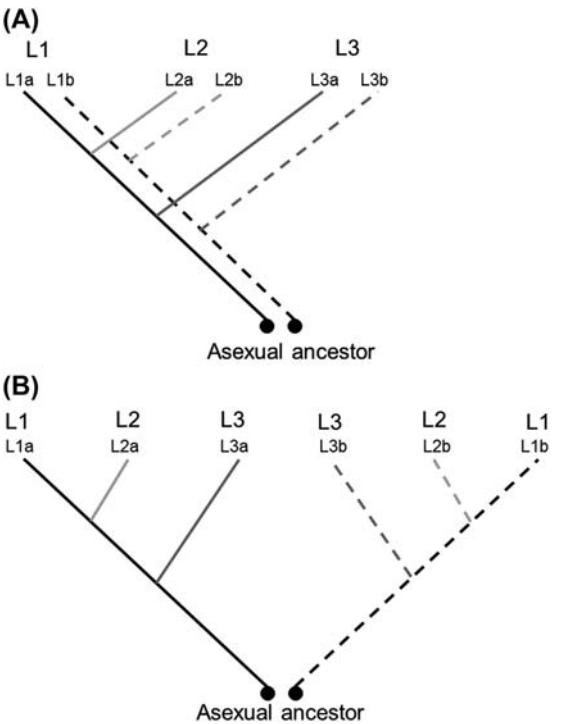


Figure 5.1 Illustration of the Meselson effect. In (A), the evolutionary relationships among three asexual diploid lineages are represented (L1−L3). The genetic divergence is also represented with varying colors providing the two alleles present in each taxon (alleles a and b). If we develop the tree corresponding to all DNA sequences (all alleles) as in (B), it is easily seen that the maximum divergence is obtained between the two alleles of each individual representing each lineage. This is what is expected in ancient clones and can be used as a criterion for detecting a long absence of sex in a group of taxa (the Meselson method).

Downloaded From Vet-eBooks.com

individual will be higher than mean divergence between lineages. This is the Meselson effect.[25,26] Another way to see it is that in lineages that have stayed clonal for a sufficient amount of time, all loci will be heterozygous for all individuals. Genomic fixed heterozygosity can thus represent an unambiguous signature of full clonality. The Meselson effect has been evidenced in *T. brucei gambiense*.[44] The discovery that bdelloid rotifers are degenerate tetraploids that probably came from an ancient hybridization and might not illustrate a Meselson effect after all makes *T. brucei gambiense* the only proven example of such a phenomenon.

Another consequence of clonality, when total, is that mutation rates are lowered (e.g., Ref. 45 and references therein), because meiosis generates more errors than mitosis. This might have long-term consequences but have not been much explored so far to our knowledge.

7. Evolution and the Paradox of Sex

The paradox of sex essentially concerns parthenogenetic multicellular organisms and, as explained earlier, microbes are not concerned. This has been the subject of an impressive amount of literature and, except plant parasitic arthropods (insects, mites) and nematodes, very few animal parasites are parthenogenetic (some nematodes, gyrodactylid monogens, rare cestodes, and trematodes).[1] It would be useless to do something more than a short reminder here. Parthenogenetic females produce twice as many offspring as sexually reproducing females that need to produce half "useless" males, which themselves cannot produce eggs. This has been named the twofold cost of sex.[46] Consequently, parthenogenetic females should quickly invade the whole planet. There are several reasons why this is not so, most of which are not exclusive and probably account together for the maintenance of sex in such situations.

First of all, as mentioned earlier, the hijacking of SR for producing clonal descents is probably extremely difficult and the diversity of tricks that evolved to achieve it, sometimes through extremely (at least in appearance) odd means, can be the sign of how difficult it is to reach that point. For instance, automictic parthenogenesis with fusion of two polar bodies illustrates this last point (see Fig. 3b in Ref. 1). The rarity of emergence of parthenogenesis, apparently restricted in few lineages (but this can be misleading because of biases in the intensity of work devoted to certain groups), can thus largely be explained by such constraints. For instance it seems impossible to evolve in mammals or in birds.

Secondly, the problem only arises for populations that exclusively reproduce either sexually or parthenogenetically and for which these two morphs compete for the same resources. This might be rare. Some aphids might correspond to this, as for instance *Rhopalosipum padi*,[47] though it is not well established how similar the ecological niche of these two morphs is.

According to the red queen hypothesis,[48] pure parthenogenetic females cannot efficiently fight against the continuously evolving aggressors (parasites and predators) or victims (preys or hosts) as compared to sexual females that produce many different

Downloaded From Vet-eBooks.com

combinations of offspring at each generation.[25] This hypothesis alone has two important drawbacks. First, in pure sexuals, the best combination is lost at the next generation. Second, most populations are not that polymorphic, are often small, and thus inbred. The possible combinations created by SR might not be that diverse or new.

Muller's ratchet[49] imposes to parthenogenetic lineages an accumulation of deleterious mutations that could lead to an eventual collapse of such lineages as compared to sexual lineages where deleterious mutations are more efficiently removed. This model alone also has two drawbacks. First, it requires several generations to work efficiently, and might even be almost silent in diploids. Second, as mentioned earlier, small sexually reproducing populations might also be affected by Muller's ratchet.

Finally, as mentioned earlier and elsewhere,[50] SR may also be viewed as a resetting process that evolved to restore the best combinations, a purpose for which it indeed evolved for in the first eukaryotes. Such a view has also the advantage to explain why SR often concerns genetically related partners, hence the evolution of reproductive isolation often observed in pluricellular eukaryotes.[29]

8. Clonal Microevolution

This aspect can be tackled differently depending on what kind of genetic information we are dealing with: neutral variation, and its use as a signature of demographic events, and variation under selection.

8.1 Neutral Loci Variability in Clonal Populations (Population Genetics Structure)

Neutral variation and its distribution in time and space can be used to make useful inferences on the population biology of the targeted organisms. Under certain hypotheses, several inferences can be made with regard to population size, dispersal, and reproductive mode. Most tools were developed for sexual species but several works have made available equivalent tools for clonal populations.[2,30,31,43,45,51,52] In that case, special care must be given to how to deal with MLGs. For A cycles, complete data sets must be kept. For I cycles, it was shown that besides analyzing complete data sets, population subdivision is better assessed if only a single representative of each MLG is kept.[53,54] For S cycles, all depends on where in the cycle individuals are sampled. A strategy similar to the one used for I cycles is to be used if individuals are sampled early after the last SR event. If individuals are sampled after a substantial amount of clonal generations, then a strategy similar to the one used for A cycles is to be preferred. Nevertheless, applying both strategies and comparing the results cost little and might represent the best option.

For A cycles, if clonal reproduction is so prevalent that no perceptible signature of any SR can be noticed, then tools specific to that situation should be used for ecological inferences. This of course must take into account some basic knowledge of the population. When the population can be assumed to be strongly subdivided in

Downloaded From Vet-eBooks.com

numerous demes, it was shown that the number of migrants can be estimated through the formula[55,56]:

$$N(m+u) = -\frac{1+F_{IS}}{4F_{IS}} \tag{5.1}$$

where N is the clonal subpopulation size, m is the proportion of migrants that each subpopulation contain, u is the mutation rate, and F_{IS} is Wright's fixation index[31,57] measuring inbreeding within individuals relative to inbreeding between individuals. In that case, estimating N and m independently, even if we assume u negligible as compared to m, is not easy and will require further studies. When the population can be assumed to comprise only two subpopulations, then more precise estimates can be made[58]:

$$N = -\frac{1+F_{IS}}{8uF_{IS}} \tag{5.2}$$

and

$$m = \frac{1}{2}\left[1 - \sqrt{\frac{F_{ST}}{F_{ST} - 4uF_{IS}}}\right] \tag{5.3}$$

where F_{ST} is Wright's fixation index measuring the between individuals inbreeding within subpopulations relative to inbreeding between subsamples. It also requires knowledge of u. Finally, when subpopulations are assumed completely isolated, their clonal size can be estimated as[59]:

$$N = -\frac{1+F_{IS}}{4uF_{IS}} \tag{5.4}$$

Now if some SR influences the distribution of genetic diversity, then it is usually wiser to use classical population genetics tools[31] except for cases of extremely rare SR events where the behavior of most parameters is odd and thus where inferences can only be very general.[30] Similar advice can be given for I and S cycles if individuals studied are sampled just after SR. Finally, mutation rate can be extrapolated from the literature regarding the kind of markers used. Nevertheless, it is now known that mitotic mutations are rarer than meiotic ones. For instance, microsatellite mutation rate is more likely around 10^{-5} in clonal populations than an average $u \sim 10^{-4}$ that applies to sexual populations.[45]

8.2 Selection and Adaptation in Clonal Populations

The vast majority of mutations are neutral or deleterious.[60] Extensive study of such mutations has explained the genetic diversity in many populations and has been useful

Downloaded From Vet-eBooks.com

for inferring population parameters and histories from data as explained earlier. Yet beneficial mutations, despite their rarity, are what cause long-term adaptation and can also dramatically alter the genetic diversity at linked sites (e.g., Ref. 61 for a review). Unfortunately, our understanding of their dynamics remains poor, especially in asexual populations.

Adaptation by natural selection occurs through the spread and substitution of mutations that improve the performance of an organism and its reproductive success in a particular environment. This happens, for example, when, in a pathogen, an allele increases in frequency in the population because it confers a certain degree of resistance against a particular drug. Most early works on the dynamics of adaptation in asexual populations considered that beneficial mutations only occurred very rarely.[62,63] Under such circumstances, the rates of adaptation of asexual populations is the same (all else being equal) as that of sexual populations and depends only on the time separating the appearance of two beneficial mutations. This conventional model, known as the "periodic selection" model remained a very influential theory until the 1990s, and so despite the previous classical works of Muller[64] that clearly showed that the dynamic of adaptation in sexual and asexual populations could be very different when beneficial mutations were common.

One particularity of the dynamic of adaptation of asexual populations when beneficial mutations are common is that beneficial mutations that have arisen independently in different individuals cannot recombine and therefore have to compete for fixation. This effect is called "clonal interference."[65–67] To date, two main models of clonal interference have been proposed: (1) the one-by-one mutation model[65] and (2) the multiple mutations model.[66,67] These two models differ in how and where new beneficial mutations appear. We do not enter into the details of these models here and we advise readers to refer to cited references for more details. We simply want to stress that, under the two models, beneficial mutations enter into competition and some beneficial mutations are therefore "wasted" during the process of adaptation.[65,68–70] This leads to a slowdown in the rate of adaptation in purely asexual populations as compared to sexual populations. Note that a similar effect was described for sexual populations in the case of physically linked genes and was called the Hill–Robertson effect.[71]

Clearly, a complete picture of adaptation in asexual populations should also include the impact of deleterious mutations. They indeed play an important role in adaptation because their presence influences the fate of beneficial mutations, and consequently affects the strength of clonal interference.[72–74] It is indeed well established that deleterious mutations can cause a severe reduction in the adaptation rate, as a consequence of reducing the effective population size. The simplest situation corresponds to the case in which only beneficial mutations that occur in individuals that are mutation-free contribute to the adaptive process.

We have here mainly focused on complete clonal organisms (life cycle A with 100% clonality). As shown all along this chapter, clonal reproduction occurs under several forms and in several life cycles. Models analyzing the dynamic of adaptation under such life cycles have not been done yet but we think that, as soon as a bit of recombination occurs, the dynamic of adaptation will be similar to the one described by models dealing with the problem of interference (or Hill–Robertson effect) in

Downloaded From Vet-eBooks.com

sexual organisms. However, since pure sexuals tend to lose the most beneficial combinations built in previous generations, clonal populations with rare sex probably display much more efficient adaptive dynamics. A rare sexual event can build an "optimal" combination that will be easily and faithfully propagated by clonal reproduction. This might help understanding the formidable adaptive speed of microbes and in particular pathogenic microbes.

9. Conclusions

Clonal reproduction is as old as life itself and is widespread in the living world. Sexual recombination appeared in Eukaryota, after this group evolved mitosis (a prerequisite for meiosis), not as a propagation tool alternative to clonal reproduction, but as a repairing tool to preserve the most harmonious combinations of the numerous genes necessary to build a eukaryotic cell. Sex is totally linked to propagation only in two pluricellular lineages (Metazoa and Metabionta). Only in those complex lineages, SR can be in competition with clonal reproduction, under certain precise circumstances. Clonality is the most important propagation mode used by pathogenic agents and its genetic consequences must be understood precisely, though SR or recombination is also very important to take into account for those diseases that practice it. When SR is so rare that no signature can be found in the genetic architecture of populations, some specific patterns arise as presence of multilocus repeated genotypes and, for diploids, fixed heterozygosity. These patterns can be exploited for demographic inferences using specific tools. If SR has even a small influence, then classical tools of population genetics can be used to infer subpopulation sizes and dispersal. It is thus possible to infer population sizes and dispersal for clonal parasites with the study of variable molecular markers, which is good news as the populations of such organisms are difficult to study directly. Such information can be vital to understand the epidemiology of diseases.

Although purely sexual populations are at a theoretical advantage as compared to purely asexual lineages as regards the dynamics of adaptation, things become less clear if the most general case is taken into account. Clones with more or less rare sex (or recombination) may indeed represent an extremely efficient (and hence widespread) way to adapt to the environment. This helps explaining the speed at which pathogenic agents respond to defense mechanisms, including pharmacologically mediated ones, of their victims.

Abbreviation List

HGT	Horizontal gene transfer
HW	Hardy–Weinberg
LD	Linkage disequilibrium
LUCA	Last universal common ancestor

Downloaded From Vet-eBooks.com

MLG Multilocus genotype
SR Sexual recombination

Acknowledgments

We thank Joseph Carl Robnett Licklider and Leonard Kleinrock, MIT initiators of Internet, without which this chapter (among so many others) could not have come to light. T.D.M. and F.P. are financed by the CNRS and IRD. F.P. is also supported by ANR MGANE SEST 012 2007.

References

1. De Meeûs T, Prugnolle F, Agnew P. Asexual reproduction: genetics and evolutionary aspects. *Cell Mol Life Sci* June 2007;**64**(11):1355−72. PubMed PMID: ISI: 000247210600005.
2. De Meeûs T, Prugnolle F, Agnew P. Asexual reproduction in infectious diseases. In: Schön I, Martens K, van Dijk P, editors. *Lost sex: the evolutionary biology of parthenogenesis*. NY: Springer; 2009. p. 517−33.
3. Cavalier-Smith T. Origins of the machinery of recombination and sex. *Heredity* 2002;**88**: 125−41. PubMed PMID: WOS:000176215300007.
4. Asker SE, Jerling L. *Apomixis in plants*. Boca Raton, Florida: CRC Press Inc.; 1992.
5. Bell G. *The masterpiece of nature*. Berkeley: University of California Press; 1982.
6. Hughes RN. *A functional biology of clonal animals*. London and New York: Chapman and Hall; 1989.
7. Jackson JBC, Buss LW, Cook RE. *Population biology and evolution of clonal organisms*. New Haven: Yale University Press; 1985.
8. Otto SP, Lenormand T. Resolving the paradox of sex and recombination. *Nat Rev Genet* April 2002;**3**(4):252−61. PubMed PMID: ISI:000174739800012. English.
9. Savidan Y. Apomixis: genetics and breeding. *Plant Breed Rev* 2000;**18**:13−86.
10. Glansdorff N, Xu Y, Labedan B. The conflict between horizontal gene transfer and the safeguard of identity: origin of meiotic sexuality. *J Mol Evol* 2009;**69**(5):470−80. PubMed PMID: WOS:000272574100007.
11. Ramesh MA, Malik SB, Logsdon JM. A phylogenomic inventory of meiotic genes: evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. *Curr Biol* 2005;**15**(2): 185−91. PubMed PMID: WOS:000226858600032.
12. Solari AJ. Primitive forms of meiosis: the possible evolution of meiosis. *Biocell* 2002;**26**(1): 1−13. PubMed PMID: WOS:000176207300001.
13. Poole AM. Horizontal gene transfer and the earliest stages of the evolution of life. *Res Microbiol* 2009;**160**(7):473−80. PubMed PMID: WOS:000271845600006.
14. Luo YN, Wasserfallen A. Gene transfer systems and their applications in Archaea. *Syst Appl Microbiol* 2001;**24**(1):15−25. PubMed PMID: WOS:000168850600002.
15. Thomas CM, Nielsen KM. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* 2005;**3**:711−21.
16. Casjens S. Prophages and bacterial genomics: what have we learned so far? *Mol Microbiol* July 2003;**49**(2):277−300. PubMed PMID:12886937.

Downloaded From Vet-eBooks.com

17. De Meeûs T, Balloux F. Clonal reproduction and linkage disequilibrium in diploids: a simulation study. *Infect Genet Evol* December 2004;**4**(4):345−51. PubMed PMID: 15374532.
18. De Meeûs T, Guégan JF, Teriokhin AT. MultiTest V.1.2, a program to binomially combine independent tests and performance comparison with other related methods on proportional data. *BMC Bioinforma* 2009;**10**(1):443. http://dx.doi.org/10.1186/1471-2105-10-443. PubMed PMID.
19. Prugnolle F, De Meeûs T. Apparent high recombination rates in clonal parasitic organisms due to inappropriate sampling design. *Heredity* February 2010;**104**(2):135−40. PubMed PMID:19812614.
20. Copley SD, Smith E, Morowitz HJ. The origin of the RNA world: co-evolution of genes and metabolism. *Bioorg Chem* December 2007;**35**(6):430−43. PubMed PMID:17897696.
21. Glansdorff N, Xu Y, Labedan B. The origin of life and the last universal common ancestor: do we need a change of perspective? *Res Microbiol* 2009;**160**(7):522−8. PubMed PMID: WOS:000271845600012.
22. Vaneechoutte M, Fani R. From the primordial soup to the latest universal common ancestor. *Res Microbiol* 2009;**160**(7):437−40. PubMed PMID: WOS:000271845600001.
23. Gargaud M, Martin H, López-García P, Montmerle T, Pascal R. *Le Soleil, La Terre...La Vie: La Quête des Origines*. Paris: Belin-Pour la Science; 2009. p. 306.
24. Gribaldo S, Brochier C. Phylogeny of prokaryotes: does it exist and why should we care? *Res Microbiol* 2009;**160**(7):513−21. PubMed PMID: WOS:000271845600011.
25. Judson OP, Normark BB. Ancient asexual scandals. *Trends Ecol Evol* February 1996;**11**(2): 41−6. PubMed PMID: ISI:A1996TT85400006. English.
26. Mark Welch DB, Meselson M. Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science* May 19, 2000;**288**(5469):1211−5. PubMed PMID: ISI:000087112600043. English.
27. Rougeron V, De Meeûs T, Kako Ouraga S, Hide M, Bañuls AL. "Everything you always wanted to know about sex (but were afraid to ask)" in *Leishmania* after two decades of laboratory and field analyses. *PLoS Pathog* 2010;**6**(8):e1001004. PubMed PMID: 20808896. Pubmed Central PMCID:2924324.
28. De Meeûs T, Renaud F. Parasites within the new phylogeny of eukaryotes. *Trends Parasitol* June 2002;**18**(6):247−51. PubMed PMID: ISI:000175784600004. English.
29. De Meeûs T, Durand P, Renaud F. Species concepts: what for? *Trends Parasitol* October 2003;**19**(10):425−7. PubMed PMID: ISI:000186058400001. English.
30. De Meeûs T, Lehmann L, Balloux F. Molecular epidemiology of clonal diploids: a quick overview and a short DIY (do it yourself) notice. *Infect Genet Evol* March 2006;**6**(2): 163−70. PubMed PMID: ISI:000236245400011. English.
31. De Meeûs T, McCoy KD, Prugnolle F, Chevillon C, Durand P, Hurtrez-Boussès S, et al. Population genetics and molecular epidemiology or how to "débusquer la bête". *Infect Genet Evol* March 2007;**7**(2):308−32. PubMed PMID: ISI:000244351400024.
32. Tibayrenc M. Population genetics of parasitic protozoa and other micro-organisms. *Adv Parasitol* 1995;**36**:47−115.
33. Tibayrenc M. Genetic epidemiology of parasitic protozoa and other infectious agents: the need for an integrated approach. *Int J Parasitol* January 1998;**28**(1):85−104. PubMed PMID: ISI:000072005200009. English.
34. Tibayrenc M. Toward an integrated genetic epidemiology of parasitic protozoa and other pathogens. *Annu Rev Genet* 1999;**33**:449−77. PubMed PMID: ISI:000084956600014. English.

Downloaded From Vet-eBooks.com

35. Tibayrenc M, Ayala FJ. The clonal theory of parasitic protozoa: 12 years on. *Trends Parasitol* September 2002;**18**(9):405−10. PubMed PMID:12377258.
36. Tibayrenc M, Kjellberg F, Arnaud J, Oury B, Brenière SF, Dardé ML, et al. Are eukaryotic microorganisms clonal or sexual? A population genetics vantage. *Proc Natl Acad Sci USA* 1991;**88**(5129−5133):5129−33.
37. Tibayrenc M, Kjellberg F, Ayala FJ. A clonal theory of parasitic protozoa: the population structures of *Entamoeba, Giardia, Leishmania, Naegleria, Plasmodium, Trichomonas*, and *Trypanosoma* and their medical and taxonomical consequences. *Proc Natl Acad Sci USA* 1990;**87**:2414−8.
38. Taylor JW, Geiser DM, Burt A, Koufopanou V. The evolutionary biology and population genetics underlying fungal strain typing. *Clin Microbiol Rev* January 1999;**12**(1):126−46. PubMed PMID: ISI:000078057000008.
39. Suomalainen E, Saura A, Lokki J. Evolution of parthenogenetic insects. *Evol Biol* 1976;**9**: 209−57.
40. Carvalho GC. Genetics of aquatic clonal organisms. In: Beaumont A, editor. *Genetics and evolution of aquatic organisms*. London: Chapman & Hall; 1994. p. 291−319.
41. Milgroom MG. Recombination and the multilocus structure of fungal populations. *Annu Rev Phytopathol* 1996;**34**:457−77. PubMed PMID:15012552.
42. Maynard-Smith J, Smith NH, Orourke M, Spratt BG. How clonal are bacteria? *Proc Natl Acad Sci USA* May 1993;**90**(10):4384−8. PubMed PMID: ISI:A1993LC72000014.
43. Halkett F, Simon JC, Balloux F. Tackling the population genetics of clonal and partially clonal organisms. *Trends Ecol Evol* April 2005;**20**(4):194−201. PubMed PMID: 000228285800010. English.
44. Weir W, Capewell P, Foth B, Clucas C, Pountain A, Steketee P, et al. Population genomics reveals the origin and asexual evolution of human infective trypanosomes. *eLife* 2016;**5**. e11473.
45. Sere M, Kaboré J, Jamonneau V, Belem AMG, Ayala FJ, De Meeûs T. Null allele, allelic dropouts or rare sex detection in clonal organisms: simulations and application to real data sets of pathogenic microbes. *Parasit Vect* 2014;**7**:331.
46. Hurst LD, Peck JR. Recent advances in understanding of the evolution and maintenance of sex. *Trends Ecol Evol* February 1996;**11**(2):46−52. PubMed PMID: ISI:A1996TT85400007. English.
47. Delmotte F, Leterme N, Gauthier JP, Rispe C, Simon JC. Genetic architecture of sexual and asexual populations of the aphid *Rhopalosiphum padi* based on allozyme and microsatellite markers. *Mol Ecol* 2002;**11**:711−23.
48. Van Valen L. A new evolutionary law. *Evol Theory* 1973;**1**:1−30.
49. Kondrashov AS. Classification of hypotheses on the advantage of amphimixis. *J Hered* 1993;**84**:372−87.
50. Schaefer I, Domes K, Heethoff M, Schneider K, Schon I, Norton RA, et al. No evidence for the 'Meselson effect' in parthenogenetic oribatid mites (Oribatida, Acari). *J Evol Biol* January 2006;**19**(1):184−93. PubMed PMID: ISI:000234543600020. English.
51. Arnaud-Haond S, Duarte CM, Alberto F, Serrão EA. Standardizing methods to address clonality in population studies. *Mol Ecol* 2007;**16**:5115−39.
52. Becheler R, Diekmann O, Hily C, Moalic Y, Arnaud-Haond S. The concept of population in clonal organisms: mosaics of temporally colonized patches are forming highly diverse meadows of *Zostera marina* in Brittany. *Mol Ecol* June 1, 2010;**19**(12):2394−23407. PubMed PMID:20465589.
53. Caillaud D, Prugnolle F, Durand P, Théron A, De Meeûs T. Host sex and parasite genetic diversity. *Microbes Infect* August 2006;**8**(9−10):2477−83. PubMed PMID: ISI: 000241992800019. English.

Downloaded From Vet-eBooks.com

54. Prugnolle F, Théron A, Pointier JP, Jabbour-Zahab R, Jarne P, Durand P, et al. Dispersal in a parasitic worm and its two hosts: consequence for local adaptation. *Evolution* February 2005;**59**(2):296−303. PubMed PMID: ISI:000227468700005. English.
55. De Meeûs T, Balloux F. F-statistics of clonal diploids structured in numerous demes. *Mol Ecol* August 2005;**14**(9):2695−702. PubMed PMID: ISI:000230573600007. English.
56. Nébavi F, Ayala FJ, Renaud F, Bertout S, Eholié S, Moussa K, et al. Clonal population structure and genetic diversity of *Candida albicans* in AIDS patients from Abidjan (Côte d'Ivoire). *Proc Natl Acad Sci USA* March 2006;**103**(10):3663−8. PubMed PMID: ISI: 000236225300031.
57. Wright S. The interpretation of population structure by F-statistics with special regard to system of mating. *Evolution* 1965;**19**:395−420.
58. Koffi M, De Meeûs T, Bucheton B, Solano P, Camara M, Kaba D, et al. Population genetics of *Trypanosoma brucei gambiense*, the agent of sleeping sickness in Western Africa. *Proc Natl Acad Sci USA* 2009;**106**(1):209−14. PubMed PMID: WOS:000262263900040.
59. Simo G, Njiokou F, Tume C, Lueong S, De Meeûs T, Cuny G, et al. Population genetic structure of Central African *Trypanosoma brucei gambiense* isolates using microsatellite DNA markers. *Infect Genet Evol* 2010;**10**(1):68−76.
60. Loewe L, Hill WG. The population genetics of mutations: good, bad and indifferent. *Philos Trans R Soc Lond B Biol Sci* April 27, 2010;**365**(1544):1153−67. PubMed PMID: 20308090.
61. Nielsen R. Molecular signatures of natural selection. *Annu Rev Genet* 2005;**39**:197−218. PubMed PMID:16285858.
62. Atwood KC, Schneider LK, Ryan FJ. Periodic selection in *Escherichia coli*. *Proc Natl Acad Sci USA* March 1951;**37**(3):146−55. PubMed PMID:14808170.
63. Atwood KC, Schneider LK, Ryan FJ. Selective mechanisms in bacteria. *Cold Spring Harb Symp Quant Biol* 1951;**16**:345−55. PubMed PMID:14942749.
64. Muller HJ. Some genetic aspects of sex. *Am Nat* 1932;**66**:118−38.
65. Gerrish PJ, Lenski RE. The fate of competing beneficial mutations in an asexual population. *Genetica* 1998;**102**−**103**(1−6):127−44. PubMed PMID:9720276.
66. Desai MM, Fisher DS. Beneficial mutation selection balance and the effect of linkage on positive selection. *Genetics* July 2007;**176**(3):1759−98. PubMed PMID:17483432.
67. Desai MM, Fisher DS, Murray AW. The speed of evolution and maintenance of variation in asexual populations. *Curr Biol* March 6, 2007;**17**(5):385−94. PubMed PMID:17331728.
68. Wilke CO. The speed of adaptation in large asexual populations. *Genetics* August 2004; **167**(4):2045−53. PubMed PMID: ISI:000223720300045.
69. Gerrish PJ. The rhythm of microbial adaptation. *Nature* 2001;**413**:299−302.
70. Rozen DE, de Visser JAGM, Gerrish PJ. Fitness effects of fixed beneficial mutations in microbial populations. *Curr Biol* 2002;**12**:1040−5.
71. Hill WG, Robertson A. The effect of linkage on limits to artificial selection. *Genet Res* December 1966;**8**(3):269−94. PubMed PMID:5980116.
72. Felsenstein J. The evolutionary advantage of recombination. *Genetics* October 1974;**78**(2): 737−56. PubMed PMID:4448362.
73. Charlesworth B. The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genet Res* June 1994;**63**(3):213−27. PubMed PMID: ISI: A1994PA73500006. English.
74. Bachtrog D, Gordo I. Adaptive evolution of asexual populations under Muller's ratchet. *Evolution* July 2004;**58**(7):1403−13. PubMed PMID: ISI:000222982800001. English.

Downloaded From Vet-eBooks.com

This page intentionally left blank

Downloaded From Vet-eBooks.com

Coevolution of Host and Pathogen

6

A.D. Morgan [1], B. Koskella [2]
[1]University of Edinburgh, Edinburgh, United Kingdom; [2]University of California, Berkeley, CA, United States

1. Coevolution of Host and Pathogen

1.1 Introduction to Coevolution of Host and Pathogen

No species is an island: every individual organism is in constant interaction with other species around it, whether it is with prey, predators, herbivores, competitors, mutualists, pollinators, or pathogens. These biotic interactions often have large effects on individual fitness and can significantly alter the evolutionary trajectory of a population. Importantly, selection imposed by species interactions can drive genetic divergence between populations and maintain diversity both locally[1–3] and globally.[4–7] This is because a given genotype might have a very different fitness in the context of one environment/community than another, as the species and genotypes with which it will interact in each environment/community are likely to differ. When biotic interactions drive reciprocal change in both populations, as one species imposes selection on the other and vice versa, the species are said to be coevolving.[8]

Coevolutionary dynamics between hosts and pathogens have been perhaps the most well-studied interspecific interaction. This is due to the tight coupling of the two players and the implications of these dynamics for understanding the structure of communities,[9] population dynamics,[10] the maintenance of sexual recombination,[11] and the trajectory of species invasions.[12] Recent research on host–parasite interactions has indicated that coevolution occurs in relatively short time periods[13–17] and that the trajectories of coevolution are strongly influenced by the spatial structure of populations.[5,18–20] For those host–pathogen interactions in which there is an underlying genetic basis to infection, both the size and genetic make-up of the pathogen population at any point in time will be a function of the frequency, and in many cases density, of susceptible host genotypes in previous generations. Similarly, the probability that a given host will become infected is a function of the frequency of pathogen genotypes in the population that can infect it, which is again determined by past genotype frequencies in both populations. Accordingly, each population acts as a moving target for the other, and it is these dynamic changes of one population in response to another that can maintain polymorphism over time, as different alleles will be favored in one generation relative to the next.[1,21,22]

In this chapter, we first discuss the process of host–pathogen coevolution. We then outline common methods for examining pathogen adaptation to hosts, and host

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00006-8
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

response to pathogens and highlight a few key examples to illustrate that this process is both common in nature and critically important in explaining the amount of genetic variation found on the planet. Finally, we discuss the implications of coevolution and summarize the importance of studying coevolution.

1.2 Antagonistic Coevolution

Pathogens, by definition, have deleterious fitness effects on their hosts and thus have the capacity to act as major selective forces on host populations. At the same time, pathogens are often reliant on their hosts for some stage of their life cycle, and so any change in the host population will have strong effects on the pathogen population. This interaction between host and pathogen will have different outcomes depending on factors ranging from the degree of pathogen specialization to the abiotic environment in which the interaction occurs. The interaction is not always a coevolutionary one; in some cases, selection only acts on one partner. For example, a generalist pathogen may sweep through a small population of a rare host species and significantly alter the host dynamics without being changed itself. Evidence from associations between Arabidopsis host populations and the generalist pathogen, *Pseudomonas syringae* suggests that the pathogen is maladapted to this host, a result that may be explained by decreased selection on the parasite population to infect this relatively ephemeral host.[23] However, due to the tight genetic interaction between many hosts and pathogens, an evolutionary change in one partner is likely to cause evolutionary change in the other, leading to ongoing coevolution. Therefore, a general definition of host–pathogen coevolution is *the reciprocal evolution of interacting hosts and pathogens*.

Host–pathogen coevolution is usually antagonistic, since an increase in fitness of one player typically leads to a decrease in fitness of the other. For example, hosts may evolve resistance (incurring higher fitness in the face of harmful pathogens) and pathogens may evolve counter infectivity. Such antagonistic coevolution may be either directional or cyclical (see Box 6.1). If it is directional, hosts and parasites evolve ever mounting resistance and infectivity in the form of an "arms-race": where future types are more resistant and infective than their ancestors.[13,24,25] This type of coevolution is typical of the interaction between bacteria and bacteriophage, and plants and their pathogens. In its simplest form, this type of directional, arms-race coevolution will lead to the extinction of one player or the other, as genetic variation is ultimately exhausted. However, in cases where there are significant costs to resistance and infectivity, these dynamics can be continuous and cyclical, as costs build up, and the "arms-race" crashes.[26] An example of such a crash is the modification of a host-cell receptor to stop a pathogen binding. The modification of the receptor may have some negative effect on the function of the receptor, and therefore affect the fitness of the host organism. The pathogen would have increasing costs associated with a reduction in its ability to bind to the receptor and so fewer successful infections. The modification of the host receptor may continue up to a point where the negative fitness effects would be so great that sensitive hosts with fully functioning receptors would be fitter than the resistant host. The cycle would then restart.[26] Although it

Downloaded From Vet-eBooks.com

Box 6.1 Infection Genetics

One critical determinate of host–pathogen coevolutionary dynamics is the underlying genetic interaction between them. Theoretical work has shown that tight genetic specificity for infection can lead to oscillations in genotype frequencies: that is, Red Queen dynamics, and the long-term maintenance of genetic diversity.[39,40] These oscillatory dynamics are key to many central theories regarding host–parasite coevolution, including both local adaptation (LA) and the maintenance of sexual reproduction.[34,36,41,42] Two models describing infection specificity in host–parasite interactions have been highly supported, although numerous others exist.

The first model is the matching alleles model (MAM); based upon a system of self/nonself-recognition molecules, where hosts can successfully defend against any parasite genotype that does not match their own.[34,43,44] A parasite must specifically match host alleles at infection loci in order for it to evade detection by the immune system and successfully infect the host. This model is typical of many invertebrate immune systems. The MAM assumes that one parasite genotype will have a different subset of susceptible hosts than another parasite genotype such that infection success is determined by both host and parasite genotype. The tight specificity leads to cyclical "Red Queen" dynamics. Evidence for this model has come from work on *Daphnia magna* and its parasitic bacterium *Pasteuria ramose*, where parasite attachment rates were used to identify a high level of parasite specificity across host genotypes.[45]

The second model, referred to as the "gene-for-gene model" (GFGM), predicts that the interaction between parasite virulence loci and host resistance loci determines successful infection.[46] The GFGM is based on resistance and virulence genes found in plants and their pathogens, respectively, and is characterized by directional "arms-race" dynamics.[24–26] At an interacting locus, pathogens can have either an avirulence or virulence gene, and the host will have either a susceptible or resistance gene. A pathogen with an avirulence gene at an interacting locus can infect a host with a susceptible gene, but not a host with a resistance gene. A pathogen with a virulence gene can infect a host with either a susceptible or resistance gene. There may be several loci involved the interaction, so initially at the start of a coevolutionary interaction a parasite may have several avirulence genes and the host has entirely susceptible genes. The host would evolve resistance at one locus and the parasite would subsequently gain a virulence gene. This process would continue at other loci until, in the absence of costs associated with infectivity and resistance, parasites become supergeneralists, infecting a wider and wider range of host genotypes, and hosts become generally resistant to wider and wider range of parasite genotypes.[13,24–26]

The MAM and the GFGM are probably two ends of a spectrum, and the interaction between most hosts and pathogens is likely to lie somewhere between the two extremes with some degree of specialization and some generalization. This may be due to costs in the GFGM. Under the GFGM gaining, several virulence

Continued

Downloaded From Vet-eBooks.com

Box 6.1 Infection Genetics—cont'd

and resistance genes may be costly to the parasite or host, which may prevent supergeneralists fixing in the population with virulence or resistance genes at every locus.[26] The cost would give a fitness advantage to a host with susceptibility genes in the presence of a pathogen with the corresponding virulence loci. Once hosts with susceptible genes increase in frequency, selection will favor pathogens with avirulence genes, as these can also infect the common susceptible hosts, but do not carry any costs associated with virulence. This will lead to cyclical dynamics like those seen in the MAM.[26] Models since 2000 have suggested that a combination of the two models might capture more biological realism and relax the assumptions required for the maintenance of genetic diversity by parasites.[47–49] For example, in one model, the pathogens have full infectivity on matching genotypes, as assumed under the MAM, but there is a continuum where other genotypes can be infected as under the GFGM except the parasites have lower fitness and the host suffers less than they would if the genotypes fully matched.[47] In this model, any departure from a pure GFGM led to cyclical dynamics, as under the MAM.

Understanding how successful infection is determined at the genotypic level is critical in understanding how disease spreads through a population. Specifically, if infection success is based solely on host resistance or parasite virulence, as is true under the strict GFGM, virulent parasites should quickly sweep through any susceptible host populations and infect most of the host population.[26] Alternatively, if infection success is determined by an interaction between host and parasite genotype, as is true under the MA model, only a subset of host genotypes will be infected, and only a subset of parasite genotypes will be infective at any given time. Testing the underlying assumption of tight genetic specificity for infection has thus far produced mixed results. Although it is clear that there exists a great deal of natural variation in host resistance and parasite infectivity,[50–54] it is less clear whether specific host–genotype by parasite–genotype interactions typically govern the outcome of infection.[55]

Evidence from natural populations of hosts and parasites has shown that invertebrate host resistance is often highly specific to parasite genotype.[56–59] However, results from experiments in which parasite specificity is selected upon via experimental passaging on single host genotypes have produced mixed results. For example, when the trypanosome parasite, *Crithidia bombi*, was passaged through individuals from a colony of worker bees, the parasite did not gain infectivity on its own colony but did lose infectivity to other, allopatric colonies.[60] Similar results were found when an RNA bacteriophage was passaged through novel genotypes of bacterial hosts.[61] It is therefore becoming clear that increased specificity is not always indicative of genotype-by-genotype interactions. When a bacterial parasite, *Holospora undulate*, was passaged on host lines of the protozoan host *Paramecium caudatum*, for example, no host line by parasite–line interactions were found despite evidence for increased infection success on sympatric host–parasite combinations.[62]

Downloaded From Vet-eBooks.com

remains unclear how ubiquitous these costs to resistance and infectivity might be, there is strong evidence that host resistance is costly from at least a few studies[27–30] and that parasite virulence is costly from a few others.[31–33]

Cyclical coevolution, on the other hand, occurs when successful infection of a host requires specific genotypic matching of pathogens. For example, host A is susceptible to pathogen A but not pathogen B, and host B is susceptible to pathogen B but not pathogen A (Box 6.1). Under this scenario, resistance and infectivity do not increase through time, as no parasite is universally virulent and no host is inherently more resistant than another. Instead, all fitnesses are determined by the frequency of "matching" genotypes in the population. Under this scenario, pathogens will evolve to infect the most common host genotype, giving rare hosts an advantage.[22,34,35] These rare genotypes might increase in frequency until they become common and eventually the target of local pathogens. These cyclical dynamics are often referred to as "Red Queen" dynamics[36] after the character in Lewis Carroll's *Through the Looking Glass* who explains to Alice that, in Wonderland, "it takes all the running you can do, to keep in the same place."[37] Similarly, populations of hosts and pathogens are engaged in a constant coevolutionary battle but are, on average, maintaining the same fitness with respect to one other. The Red Queen metaphor is also used more generally to describe antagonistic coevolution whether dynamics are cyclical or directional.[38]

1.3 The Evolution of Pathogen Virulence

Although it is intuitively clear that pathogens might harm their hosts as a by-product of passing themselves on from one generation to the next (e.g., by redirecting host resources away from host reproduction and into pathogen reproduction), it is less clear why there are more virulent pathogens that kill or sterilize their hosts. The dilemma arises because an increase in pathogen fitness, via greater within-host reproduction, might lead to a decrease in fitness via lower rates of transmission, if the host becomes too sick to interact with other hosts or spread infectious propagules into the environment. This "trade-off hypothesis" is the most popular evolutionary explanation for why pathogens often do not reach their maximum reproductive potential.[63–65] Increases in virulence can accompany shifts to new host populations or species,[66] drastic changes in host population size or structure,[67] or competition with other pathogens.[68,69] However, ongoing coevolution between host and parasite populations is expected to lead to decreased virulence, as fitness of both populations is optimized.

Evidence for decreased virulence over time has been demonstrated in experimental populations of Red Flour Beetles, *Tribolium castaneum*, and the microsporidian parasite, *Nosema whitei*. After only 11 generations of experimental coevolution, parasite lines became less virulent, as measured by host mortality, without losing their ability to infect hosts.[70] Further evidence comes from experimental systems of bacteria and plasmids (circular strands of DNA often carried by bacteria that can carry beneficial genes, such as those conferring antibiotic resistance). Plasmids can be considered parasitic in that hosts harboring these elements suffer a reduction in growth rate, possibly due to the additional expression of plasmid products, which competes for the host ribosomes with the expression of host genes.[71] In an

Downloaded From Vet-eBooks.com

experimental study,[72] the costs of carrying a plasmid were reduced during experimental evolution, albeit via changes in the host only. A different study demonstrated that genetic changes in both the host cell and the plasmid lead to increases in reproductive fitness of the host cell.[73]

Aside from the trade-off model, there have been several other theories to explain the evolution of virulence. Some have suggested that mixed infections of different pathogen genotypes within a single host may have important effects upon virulence, in some cases decreasing virulence while in others increasing it.[64,74] Models where virulence increases have a similar assumption to the trade-off model: selection will favor the parasite with the fastest within-host growth rate, rather than a more prudent host exploiter. The parasite with the faster growth rate is predicted to outcompete the slower growing parasite and to have a higher probability of transmission, leading to the evolution of higher virulence than that expected for single infections. Alternatively, if parasites produce a "public good" that are utilized by all the parasites within a host, mixed infections may select for cheating behavior because of low relatedness (i.e., they are different genotypes) between parasites.[75] Examples include siderophores,[76] which are iron-scavenging molecules in bacteria, and coat proteins in viruses.[75] Such molecules may be costly for a parasite to produce. If a parasite "cheats" and does not produce them, but uses the molecules produced by a competing parasite, it does not pay the costs but gains the benefits, giving it a higher growth rate or competitive advantage. Such cheating behavior will therefore have a selective advantage, and the cheats will increase in frequency.[76] However, if there are too many cheats, there will not be enough parasites producing the "public goods" to support all the cheats, decreasing the growth rate of the parasite population, and ultimately its virulence.[77]

At an even greater extreme, initially parasitic organisms may evolve to benefit the host by *increasing* the host's fitness, changing the interaction to a mutualistic one.[78] There is evidence for this type of transition between a grain weevil, *Sitophilis zeamais*, and a bacterial mutualist, *S. zeamais* primary endosymbiont (SZPE). The genome of SZPE encodes a type III secretion system, and expression of these genes coincides with the timing of bacteriome infection within a developing weevil.[79] It is likely that the ancestor of SZPE was originally pathogenic, as type-III secretion systems are found in a diverse range of bacteria pathogenic to plants or animals, including *Salmonella* spp. and *Pseudomonas* spp.[80] and are used by these pathogens to invade the host cell.[81] It is likely that through the course of evolution, SZPE has evolved to become a mutualist, but still uses the same method to enter its host as its ancestral pathogenic bacterium did.[79]

It is important to note that virulence is not necessarily a fixed characteristic of a pathogen. Rather, virulence is often context dependent and can be influenced by host condition,[82,83] host density,[84,85] or interactions with species at other trophic levels.[86] Understanding the evolution of virulence is critical to understanding the process of host–pathogen coevolution because the magnitude of parasite-mediated selection on host populations is a direct function of both pathogen prevalence, which determines the likelihood of becoming infected; and pathogen virulence, which determines the fitness cost of being infected.

Downloaded From Vet-eBooks.com

2. The Process of Antagonistic Coevolution

2.1 Introduction to the Process of Antagonistic Coevolution

There are several factors that are thought to affect the dynamics of antagonistic coevolution, including both biotic and abiotic factors. The biotic factors include the genetic basis of host–pathogen interactions,[26,34] mutation and recombination rates,[87] generation time,[87] and interactions both with other parasites and with the host microbiome.[88–90] Abiotic factors include environmental productivity and barriers to gene flow. Other factors, such as migration rate, may be a combination of biotic and abiotic effects. Together, these factors may affect the mode and tempo of coevolution,[91] or may give either the host or pathogen an evolutionary advantage over the other. When either the host or parasite population has an evolutionary advantage over the other, it can rapidly adapt to changes in its local coevolving partner. Theory predicts that the parasite will, more often than not, have the evolutionary advantage over the host due to its typically higher migration and mutation rates, which increase the genetic variation on which selection can act, and faster generation times, which increase the speed of selection.[87,92,93]

2.2 Migration, Mutation, and Recombination

The supply of new genetic diversity plays a crucial role in shaping coevolution. For hosts and pathogens to coevolve, there needs to be a constant input of new alleles upon which selection can act as one population responds to changes in the other. Genetic diversity may be increased by mutation, recombination, or migration rates, all of which can be affected by population size. Mutation and recombination have the potential to generate novel genetic diversity within a population. Migration can also introduce novel alleles if there is spatial structuring. For example, populations are often thought to exist as metapopulations (populations divided into discrete subpopulations), resulting from environmental factors such as differences in productivity or geographic barriers. Coevolution may then drive divergence between subpopulations, as they follow different coevolutionary trajectories.[4,94,95] Low rates of migration will introduce variation from one subpopulation to another, but high rates of migration might decrease genetic diversity as the metapopulation becomes homogenized. Population size is also related to diversity, but indirectly. A large population will have a higher total number of mutants and migrants than a smaller population, when the mutation and migration rates are equal; and it will also reduce the chances of beneficial mutations being lost by drift.[96]

If mutation, recombination, migration rates, and population sizes are equivalent between hosts and parasites, then they are predicted to coevolve together at similar rates. An increase in any of these factors for both coevolving organisms is predicted to increase the rate of coevolution, as they will increase the genetic supply rate, shortening the time for reciprocal adaptation to occur. It is more likely, however, that these factors will differ between host and parasite populations, giving one of the coevolving partners an evolutionary advantage. Since parasites typically have higher mutation,

Downloaded From Vet-eBooks.com

recombination, migration rates, and larger population sizes, they can rapidly respond to changes in local host populations and are predicted to be ahead in the coevolutionary race.[10,92,93,96,97]

2.3 Generation Time

Generation time is also thought to be an important determinant of rate and strength of coevolution. A shorter generation time allows favorable genotypes that have arrived in the population by mutation, recombination, or migration to rapidly increase in frequency.[87] In most cases, parasites have shorter generation times than their hosts. Although conventional wisdom suggests that the coevolving partner with the fastest generation time gains an evolutionary advantage, theoretical predictions, and empirical data suggest that this may not always be the case.[87,97] A faster generation may allow an organism to become rapidly adapted to host, but this may come at a cost of purging the genetic diversity of a population, if the supply of new genetic diversity is limited by low mutation, migration, or recombination rates. If the host subsequently adapts to the parasite, the parasite is less able to counteradapt, due to its low genetic diversity.

2.4 Environmental and Community Context

In addition to the factors influencing the rate of population change outlined earlier, the trajectory and outcome of host−pathogen coevolution will be strongly influenced by both the community context and the abiotic environment in which it occurs. The geographical mosaic theory states that coevolution is shaped by three genetic and ecological attributes of species interactions: coevolutionary hot spots and cold spots, whereby the intensity of reciprocal selection among populations differs; selection mosaics, whereby the structure of the interaction differs among environments; and remixing of coevolved traits, whereby gene flow, mutation, genetic drift, and local extinction result in a continual reshuffling of coevolved genes among populations.[5,98] This geographic variation can result from genetic divergence among populations and/or by differing abiotic or biotic environments.

Among the more obvious examples of biotic factors that might alter the outcome of coevolution across a geographic mosaic are (1) the presence of alternate host species for more generalist pathogens, (2) the prevalence of other parasite species within a community, and (3) the presence or absence of final host species for parasites with complex life cycles or hyperparasites (i.e., parasites that infect parasites). For example, coevolution between polyphagous insects and their parasites is likely to be influenced by the plant upon which the insect feeds. The plant environments may differ in regard to chemistry, architecture, or palatability; all of which could influence the fitness of hosts, fitness of parasites, and the interaction between them (reviewed in Cory and Myers[99]). Host plant environment has also been shown to influence the infectivity, virulence, and transmission probability of nucleopolyhedrovirus among island populations of western tent caterpillars, *Malacosoma californicum pluviale*.[100] A similar result was found for the interaction between protozoan parasites, *Ophryocystis elektroscirrha*, and monarch butterflies, *Danaus plexippus* L. across two milkweed species.[86]

Downloaded From Vet-eBooks.com

Variation in host plants is also likely to influence coevolution between bacterial pathogens and hyperparasites, such as bacteriophage. For example, a study of phage adaptation to natural populations of *P. syringae* on horse chestnut trees suggested that the microenvironment within the tree host (surface vs. interior of leaves) determined the magnitude of phage adaptation to local bacteria.[101] For parasites with multiple hosts, access to hosts can differ across space and therefore alter the coevolutionary potential of interactions. An elegant example of this comes from a study of trematodes infecting both the New Zealand mud snail, *Potamopyrgus antipodarum* and dabbling ducks. In this case, coevolution between the parasite and its intermediate host, the snail, has been shown to be disrupted in deep habitats within the lakes, as the parasite is unlikely to reach the duck final host.[102] These studies emphasize that the biotic environment, in addition to the abiotic environment, can create selection mosaics across space.[5]

2.5 The Influence of the Microbiome on Host–Pathogen Interactions

The role of the microbiome in shaping susceptibility to pathogens has been the focus of considerable work[103–105] including building empirical evidence across systems.[106,107] Interactions between the host immune system and commensal members of the microbiome can have important consequences for host–pathogen interactions. For example, transplantation of microbiota among bumble bee hosts dramatically alters their susceptibility to the parasite *C. bombi*,[108] and the microbiota associated with the arabidopsis leaf has been shown to alter susceptibility to a fungal pathogen.[90] Evidence from the tick, *Dermacentor andersoni*, indicates that a bacterial member of the microbiota, *Rickettsia bellii*, can shape host susceptibility to *Anaplasma marginale*, a tickborne pathogen of livestock.[89] Furthermore, pathogen effectors are known to play a role in deregulating host immunity but have also been hypothesized to influence interactions between the pathogen and host microbiota,[109] so these interactions can be both direct and indirect. In humans, the cytokine IL-22, which is produced by epithelial immune cells in response to bacteria in the intestine, has been shown to play a role in pathogen resistance but is also important in shaping microbiome composition.[110] As such, it is clear that our understanding of host–pathogen coevolution and the spread of disease must take into account the microbiome, and the complex interactions between host genetics, the microbiota, and pathogen invasion will no doubt play a large role in our understanding of host–pathogen coevolution as we move forward.

2.6 The Effect of the Abiotic Environment on Coevolution

The abiotic environment can play a key role in shaping coevolutionary interactions. Studies with bacteria and phage have shown that increasing environmental productivity can increase the rate of coevolution,[111] and change coevolutionary dynamics from cyclical to arms-race dynamics.[112] This is probably due in part to each of three different reasons: (1) the population density of the bacteria and phage is higher in more productive environments, increasing the effective population size, and so the supply of new

Downloaded From Vet-eBooks.com

resistance and infectivity mutations; (2) an increase in population size increases the encounter rate and hence the strength of selection for new resistance and infectivity; and (3) in a more productive environment, there are lower costs of resistance, due to reduced competition for resources.[111,112] Coevolution in differentially productive environments has been shown to increase LA of phage to bacteria from the environment with the same productivity, probably due to the different coevolutionary dynamics dominating at different productivity levels.[113,114] Similar effects of reduced productivity on coevolution have been seen between bacteria and phage in soil, where coevolutionary dynamics were cyclical in the relatively low nutrient soil environment, compared to arm-races dynamics in standard nutrient-rich laboratory media.[115]

A further example of the abiotic environment shaping coevolutionary interaction has been shown by the effect of stressful increases in temperature on bacteria–phage coevolution.[116] Coevolution reduced phage adaptation to higher temperatures, driving them extinct, whereas phages were able to adapt to higher temperature in the absence of coevolution. Again, this is probably due to similar mechanisms at play to those involved in differences in coevolution between environments with differing productivity. Reduced phage population sizes and costs of infectivity, meant they could not "keep up" with the coevolving bacteria and simultaneously adapt to higher temperatures, leading to their demise.[116]

3. Testing for Host–Pathogen Coevolution

3.1 Introduction to Testing for Host–Pathogen Coevolution

Several different methodologies have been used to test for coevolution between hosts and pathogens. Coevolution can be directly measured through time, but to successfully do this, a system must allow for measurement of changes that have occurred through time and testing of whether these changes can be attributed to coevolution. Furthermore, the coevolutionary change must be rapid enough to be detected by the chosen methodology within the timescale of the experiment. The direct measurement of coevolution has been achieved in several different ways including the simultaneous measurement of host resistance and parasite infectivity over time and of population genetic changes. For systems in which direct testing is not feasible, due to either timescale or difficulty of experimental manipulation, evidence of coevolution can be gleaned from studies of adaptation across space by studying reciprocal adaptation of parasites and hosts from multiple populations.

3.2 Direct Comparisons Between Coevolving Organisms Across Time

Perhaps the most straightforward way to test for host–pathogen coevolution comes from experimental systems in which reciprocal changes over time can be explicitly compared. These "time-shift" experiments, as they are sometimes known[117] have been achieved in several different ways but is most commonly utilized in microbial

Downloaded From Vet-eBooks.com

systems. Here, we highlight how coevolution between bacteria and bacteriophage can be measured in the laboratory.

Microbial systems are highly amenable models for the study of coevolutionary processes.[118,119] They have large population sizes and short generation times that allow rapid coevolution in a short period of time: over a matter of days and weeks. Multiple populations can be kept in a laboratory enabling easy replication of experiments, and variables of interest can be directly manipulated whilst controlling for all other effects. Perhaps the key advantage of using microbes to study coevolution is that they can be frozen and stored in "suspended animation" at regular intervals during coevolution experiments. These frozen lines give a "living fossil record" where samples from different time points can be directly compared to show how the populations have changed over time.

The majority of these bacteria−phage studies use lytic phage that infect a given host bacterium, hijacking its cellular machinery and turning it into a "factory" that produces more phage progeny inside the cell. In order for phage to "escape" the host cell and infect other host cells, they must burst the host cell open, beginning the cycle again. Because phage are obligate killers, there is strong selection for bacteria to evolve resistance, and equally strong selection for counter adaptation by the obligatory parasitic phage to infect. Lysogenic phages have also been used as model organisms for host−pathogen evolution and are an interesting contrast in that they are not always obligate killers and are often vertically transmitted between bacterial generations. After infecting a host cell, a lysogenic phage may go down one of two paths: either producing more copies of the phage and lysing the host cell (as the lytic phage do), or integrating into the host genome and being transmitted vertically to the next generation of the bacteria. Therefore, lysogenic phage can be used as a model to investigate the processes that favor horizontal versus vertical transmission and the subsequent evolution of virulence.[120]

An example of how coevolution can be measured in a bacteria−lytic phage system is illustrated by the bacterium *Pseudomonas fluorescens* SBW25 and the lytic DNA phage SBWΦ2.[13] This system has been shown to coevolve in the laboratory for more than 500 bacterial generations.[15,121] The bacteria and phage (at least in the early stages of coevolution) typically follow a GFGM of coevolutionary interaction, where the bacteria and phage evolve to become evermore resistant and infective, respectively, to a wider range of genotypes.[13]

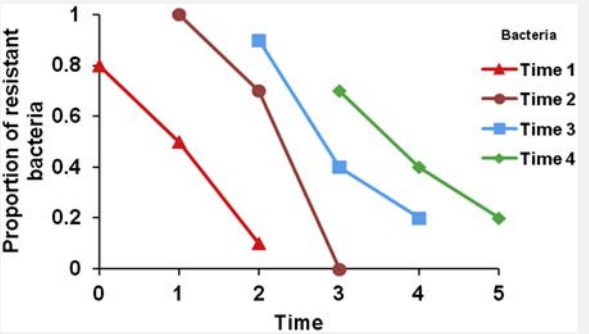
To measure coevolution during the course of an experiment, samples of bacteria and phage are frozen and stored at regular intervals. After the specified period of coevolution, bacterial colonies are isolated from each of the frozen samples. These colonies are streaked on an agar plate across samples of a population of phage isolated from either (1) a time point before the focal bacteria was isolated, (2) the same time point as the focal bacteria, or (3) a time point after the focal bacteria. After incubation, the bacterial colonies are scored as either sensitive or resistant to phage depending on ability to grow over the phage zone. As coevolution in these experiments is typically escalatory, the majority of colonies in a population are resistant to phage from previous time points, an intermediate number of colonies are resistant to phage from the same time point, and colonies are mostly susceptible to phage from later time points. This

Downloaded From Vet-eBooks.com

gives a negative change for the proportion of bacteria resistant to phage though time as shown in Box 6.2. The steepness of the slopes indicates the rate of coevolution, with steeper slopes indicating that coevolutionary change is occurring more rapidly.[13,122] This allows for a comparison between different factors that may affect the rate of coevolution, such as mutation rate.[123]

Another way to compare the rates of coevolution between populations of *P. fluorescens* and SBW25Φ2 is to measure resistance and infectivity ranges. Because

Box 6.2 Rates of Coevolution Between the Bacterium *Pseudomonas fluorescens* SBW25 and Phage SBW25Φ2



This figure shows a stylized example of the typical relationship between the resistance of bacteria to phage from different time points, thus indicating the rate of coevolution. All lines represent a single bacterial population but are broken up into time points whereby each separate symbol represents the bacteria from a corresponding time point. The three points on each line (from left to right) represent the resistance of bacteria from one particular time point to phage from the same population but from the (1) previous time point, (2) the same time point the bacteria under test were isolated from, and (3) the next time point. Therefore, the lines shows a negative slope as the infectivity of the phage has increased through time. The steepness of the slope indicates the rate of change through time and so consequently the rate of coevolution.

The graph also illustrates how bacterial resistance increases through time. Where there are two data points at the same point in time, the bacteria from two time points (two different data lines) are being compared on the same phage from one point in time, that is, the contemporary phage for bacteria from one point in time is the past phage for the bacteria from the subsequent point in time. In this graph, the bacteria always have higher resistance to a phage population from a particular time point than bacteria from the previous time point, so we can conclude that the bacteria has evolved increased resistance through time.

Downloaded From Vet-eBooks.com

coevolution is directional in the early stages and the bacteria and phage become more resistant and infective to a wider range of genotypes, they follow a predictable trajectory. Therefore, bacteria and phage from faster coevolving populations will have wider resistance and infectivity ranges, as they will be further along this trajectory than slower coevolving populations.[13] To determine resistance and infectivity ranges in the *P. fluorescens*—SBW25Φ2 system, bacterial colonies from each population are streaked across phage from the same time point but from all the different populations. This gives an average measure of resistance of the bacterial population to all the phage and the infectivity for all phage populations on the bacteria. Typically, comparisons are between different treatments where a factor that is predicted to change the rate of coevolution, such as migration rate[15] or generation time,[121] is manipulated. In this case, multiple replicates are used for each treatment, and the resistance of each replicate population of bacteria is measured against all phage replicates from all replicate populations, and all treatments.

As an alternative to examining signatures of coevolution across populations, time-shift experiments within a population allow for the direct examination of parasite and host change over time.[117] As an example of a time-shift experiment from the field, an eloquent study by Decaestecker et al.[124] took sediment cores from a pond that contained dormant eggs of the waterflea *D. magna* and dormant isolates of one of its parasites, the bacterium *Pasteuria ramosa*. These sediments contained about 39 years of coevolutionary history preserved in a sequential "living fossil record," which is effectively the same as bacteria and phage being stored in the freezer during coevolution experiments, yet over a much longer timescale. The authors examined the resistance of *Daphnia* to the parasites from one layer below (past), the same layer (contemporary), and the layer above (future). Contemporary parasites were found to be more infective than the past and future parasites, which was consistent with the MAM of Red Queen coevolution. Time-shift experiments have also been used to characterize bacteria–phage coevolution within the leaf microbiome of horse chestnut trees.[125,126]

3.3 Measuring Population Genetic Change

In a coevolving system, genotypes of hosts and pathogen will change in frequency through time, as one responds to selection imposed by the other. Molecular methods can enable the tracking of host and parasite genotypes through time and give an indication whether they change in frequency, provided that the system coevolves rapidly enough in the period of measurement. In a 2010 study with the *P. fluorescens*—SBW25Φ2 bacteria–phage system, Paterson et al.[7] allowed phage populations to experimentally evolve in the presence of either (1) coevolving bacterial host populations or (2) static, nonevolving bacterial populations, in which the bacteria were continually discarded and replaced with the ancestral strain. After 24 days, the entire genome of each phage line was sequenced. The genome of SBW25Φ2 is only about 40 kilobase pairs long, which is 100 times smaller than the genome of *E. coli*, allowing for rapid sequencing and analysis. The authors then compared the number of nonsynonymous mutations in the coevolved and evolved phage, relative to the ancestor and

Downloaded From Vet-eBooks.com

found that: coevolved phage had double the genetic divergence from the ancestor than the evolved phage; the coevolved phage had more mutated sites than the evolved phage; and there was more genetic divergence between populations of the coevolved lines than populations of the evolved phage. The results clearly show that coevolution, relative to directional evolution, leads to increased genetic divergence between populations and, ultimately, to the maintenance of genetic variation over space and time.

3.4 Pathogen-Mediated Rare Host Advantage

In 1949, Haldane suggested that parasites could be a significant evolutionary force, as they are under selection to infect the most common genotypes in a host population, thereby giving a fitness advantage to rare host genotypes. Specifically, when a host genotype is common in a population, any parasite able to infect that genotype will have a large fitness advantage, and will thus increase in frequency over time.[11,34,36,127] This will, in turn, lead to a decrease in the frequency of the susceptible host genotype and a subsequent decrease in the corresponding parasite genotype, further driving populations apart via parasite-mediated selection. Although this hypothesis has fueled a good deal of theoretical investigation,[127–130] there have been relatively few empirical tests of host rare advantage.[53,131–134] A key feature of Red Queen dynamics is the time lag between the rise in the frequency of a recently rare and resistant host genotype and the subsequent chance introduction of a matching parasite genotype via migration, mutation, or recombination. Once introduced, this parasite would realize a significant fitness advantage and, after a time lag, drive down the frequency of its host in the population. This time lag (or phase difference) is essential for driving oscillatory dynamics and has therefore been the focus of much theoretical work.[127]

Parasite-mediated negative frequency-dependent dynamics can be tested for directly, using either an experimental evolution approach or a time-shift experiment. The process can also be examined indirectly by following infection dynamics over time in natural populations. One system that has proven ideal for these methods is the New Zealand mudsnail, *P. antipodarum* and its trematode parasite, *Microphallus sp*. Upon successful infection, the trematode sterilizes, but does not kill, its snail host. Instead, the parasite reproduces within the snail and remains there, as metacercaria, until the snail is eaten by a duck, the trematode's final host. Given the parasite's high virulence (as a sterilized host has zero reproductive fitness) and thus strong potential as a selective agent, the lack of direct horizontal transmission and the relatively short generation times, this system is amenable to experimental coevolution methods in the laboratory. For example, a recent experiment, in which artificial populations of snails were evolved (1) with coevolving parasites, (2) with parasites that were lagged behind by one host generation, or (3) the absence of the trematode parasites, showed evidence for time lagged tracking of host populations by local parasites. After only six host generations, there was evidence for reciprocal change in both the host and the parasite populations, and hosts were found to be more resistant to parasites that were lagged behind than they were to coevolving parasites.[16]

This result was then followed up with a direct test of whether parasites were disproportionately infecting common host genotypes; thereby giving rare host genotypes a

Downloaded From Vet-eBooks.com

fitness advantage.[134] The genes involved in determining infection for this system are unknown, but the asexual reproductive mode of the snail means that any infection alleles will be necessarily linked to neutral, allozyme markers. By comparing genotype frequencies of each experimental population across three time points, the start, midpoint, and end of the experiment, the authors were able to demonstrate that the initially common clone declined in frequency over time in the presence, but not in the absence, of parasites. These results are consistent with negative frequency-dependent dynamics, as predicted under the matching alleles (Red Queen) model of coevolution and support previous evidence, from the field, that the trematode can impose strong selection on host populations[135] and maintain host genetic diversity over time via rare advantage.[17,136]

There have also been a number of key studies from *Daphnia* and their parasites that have directly measured the change in frequency of host and/or pathogen genotypes under both experimental[137] and field conditions.[133,138,139] *Daphnia* has been used as a model host organism to examine coevolution with a number of different naturally occurring parasites.[140] In one study, genotypic composition of natural *D. magna* populations was compared before and after epidemics of the bacterial pathogen *P. ramosa*. Resistant host genotypes were found to dominate the population after the parasite epidemic. Also, parasitism temporarily decreased genetic diversity within a population, as all susceptible genotypes were wiped out.[138] Another elegant coevolution study used the *Daphnia galeata* × *hyalina* × *cucullata* species complex to investigate negative frequency-dependent selection imposed by four of its common parasites: a protozoan, a fungal-like oomycete, and two bacterial species.[133] The authors tracked changes in host genotypes through time across natural lake populations using allozyme analysis. By comparing these changes with dynamics in populations where no infections were found, the authors show that, on average, the most common host genotype was under-infected by parasites. This indicated that the host had an evolutionary advantage, which the authors suggested could be because hosts can migrate between lakes via birds transporting their eggs, while the parasites could not. Secondly, it was found that in most cases, the common genotype declined through time in the presence of parasites, but not in the populations without parasites, clearly demonstrating parasite-mediated, negative frequency-dependent selection.

3.5 Pathogen Local Adaptation

"Red Queen" dynamics are considered to be one of the major driving forces of pathogen LA, defined as either (1) the better performance of a local parasite on its local host compared to other, allopatric parasites or (2) the better performance of a parasite on its local host compared to its performance on other, allopatric hosts.[2,15,141–143] A host genotype that is common in one population, and thus being targeted by local parasites, is unlikely to also be common in another at a given point in time. However, since parasites are lagged in their tracking of host genotypes, that is, are always responding to changes in the host population, it is predicted that parasites will occasionally be locally maladapted and thus do better on a population of allopatric hosts.[40] For many systems, it is difficult, if not impossible, to measure coevolutionary change through time.

Downloaded From Vet-eBooks.com

Therefore, it is often easier to study the outcome, consequences, or signatures of coevolution rather than the process itself. Pathogen LA, as a signature of coevolution, is relatively easy to measure, and can be examined over space (i.e., across multiple populations) as a way to understand what is likely happening over time.

The degree of parasite LA is essentially a measure of the strength of host–parasite coevolution. Parasites that are more closely able to track local host populations, and thus drive changes in local host dynamics, are expected to do better on their own hosts than they would on a randomly picked allopatric host source. The absence of LA, however, can be interpreted in one of three ways: first, that parasites are not currently successful in tracking the host population; second, that hosts are ahead in the coevolutionary game and responding to selection more effectively; or third, that coevolution is not occurring, or is weak, in that host–parasite population (i.e., that the population is acting as a coevolutionary cold spot[98]). As predicted, parasites are often found to be locally adapted to host populations (reviewed in Refs. 97,144,145), suggesting that coevolutionary dynamics are driving population divergence in many natural systems. However, there are also many cases in which parasites are found to be maladapted, indicating that hosts can often be ahead in the coevolutionary battle.

4. Implications of Coevolution

4.1 Diversification and Speciation

Host–pathogen coevolution may cause rapid divergence between populations that are isolated, or have minimal levels of migration between them.[94] Hosts and pathogens impose strong selection pressures upon each other, so host–pathogen coevolution happens relatively quickly in evolutionary time, and represents an interesting case in which ecological and evolutionary timescales might overlap. By chance, different populations will follow divergent trajectories,[94] so different hosts and parasites may dominate separate populations. In one study, coevolving populations of bacteria and bacteriophage were found to have higher allopatric diversity relative to control populations.[4] Such rapid between-population divergence is a prerequisite of LA, as populations must be different for differential performance of hosts and parasites.[10,87,96,121,141,146]

Ultimately, populations of hosts and pathogens may diverge so much that they become separate species. Although there is good evidence that host–pathogen coevolution can lead to sympatric diversification and speciation of parasites (e.g., following a host shift[147–149]), no direct evidence that host–parasite coevolution has caused host speciation exists. Several studies have shown that the phylogenies of hosts and parasites are congruent, suggesting cospeciation over time.[150–153]

4.2 The Maintenance of Genetic Diversity

The evolution and maintenance of sex is a central theoretical problem in biology. This is because there is a "cost of males," who do not produce offspring themselves.

Downloaded From Vet-eBooks.com

Therefore, an asexual female would be able to produce twice as many reproducing offspring as a sexual female, and sexual reproduction should be severely disadvantageous.[154,155] Despite the high theoretical cost of sex, most eukaryotes are still sexual. Several theories have been suggested to explain why sexual reproduction is retained. Some simply suggest that it is physically impossible to revert back to asexuality, as sex may be an integral part of the organism's development, as for example, meiosis in ciliates allows an escape from senescence[156]; or the benefits of male care outweigh the costs.[157] One of the two major explanations is that the recombination associated with sexual reproduction purges deleterious mutations and reverses "Muller's Ratchet."[158,159] The other, the Red Queen hypothesis, is that parasites play a role in the maintenance of sex by selecting for rare or novel genotypes.[42,43,49,160] Sexual recombination brings together genes from two genomes and can create or recreate rare/novel genotypes. This allows the host to constantly change every generation and "keep up" with a rapidly evolving parasite. In addition to the evidence for parasite-mediated rare advantage discussed earlier in the chapter, there is direct evidence for increased meiotic recombination within experimental populations of the red flour beetle *T. castaneum* in the presence of a parasitic microsporidian.[161]

Like sex, a high mutation rate may also introduce the required genotypic diversity to allow a host to keep up with a rapidly evolving parasite. Although an elevated mutation rate is typically disadvantageous when an organism is adapted to its environment (as deleterious mutations will outweigh beneficial mutations), it may be advantageous when an organism is in a new or changing environment.[162] For hosts, a parasite may act as a constantly changing environment, and thus a host with a higher mutation rate might benefit. This has been supported by an experimental evolution study that showed laboratory populations of bacteria were more likely to evolve a higher mutation rate, in the order of 50−100 times higher, when coevolving with phage than populations that were not exposed to phage. The mutations that conferred a higher mutation rate were in genes involved in the DNA repair pathway.[163]

5. Summary/Future Outlook

Host−pathogen coevolution is a critical and rapidly paced evolutionary force, shaping both the diversity and population structure of hosts and their pathogens. Coevolution has been demonstrated in a diverse set of host−parasite systems and, due to the ubiquity of parasites, it is likely to be very widespread across ecosystems. Although there is a large body of literature on host−pathogen coevolution, there are still several open questions in need of empirical investigation. For example, the question of what makes a pathogen more virulent, instead of mutualistic, is far from being resolved. The commonly known trade-off model of the evolution of virulence is contested by some researchers,[164] but there is little evidence supporting alternative hypotheses. Another open question is why some pathogens evolve to be specialists and others to be generalists. Again, there is thought to be a trade-off underlying the polymorphism in pathogen strategies whereby pathogen specialization allows for increased infectivity on a given host but decreased infectivity at the community level.[165,166]

Downloaded From Vet-eBooks.com

Perhaps the largest and most important open questions are regarding how human activity impacts host−pathogen coevolution. It remains to be determined whether knowledge of host−pathogen interactions can be beneficially applied to manipulate the outcome of coevolution. For example, theory and empirical evidence has shown that migration, particularly asymmetric migration of host and pathogen can radically alter host resistance and pathogen virulence.[10,15,96,167] Recent centuries have seen increased movement of humans over greater distances, especially with the advent of air travel. With them have travelled their pathogens, pathogens of animals and plants, and animal and plant hosts. How this movement has impacted on the evolution of disease of humans, and diseases in natural ecosystems and economically important animals and crops, has received little investigation so far. For example, we need to understand how parasites influence species invasions[12,168,169] and how host shifts might change parasite virulence and transmission.[61,170,171]

However, human activity could deliberately manipulate host−pathogen coevolution may shift the balance of in favor of one of the coevolving organisms. For example, when using pathogens as biocontrol agents to kill pests in agriculture, it may be advantageous to manipulate coevolution, such as increasing migration, to tilt the balance in favor of the pathogen, and away from the host pest. The medical field could thus use knowledge of coevolution to reduce the effects of disease. For example, bacteriophage have been suggested as an alternative to antibiotics in treating bacterial infections.[172,173] The advantage that bacteriophage have over antibiotics is that they can evolve to overcome resistance, whereas once a bacterium becomes resistant to an antibiotic, it is no longer of use. In this case, knowledge of the evolutionary theory behind host−pathogen coevolution could tip the evolutionary advantage towards the bacteriophage. Similarly, coevolutionary theory could be used to alter the outcome of coevolution in favor of the host in halting the spread of disease. Thus the area where knowledge and manipulation of coevolution could have its most dramatic application is within the medical field. The approach has been dubbed “Darwinian Medicine” and has received a lot of attention since the late 1990s.[174] Indeed, it has been suggested that all medics should be obliged to study evolution.[175] To parasitize the famous phrase from Dobzhansky,[176] we suggest that “no disease makes sense except in the light of coevolution.”

References

1. Haldane JBS. Disease and evolution. *Ric Sci* 1949;**19**:68−76.
2. Parker MA. Disease impact and local genetic diversity in the clonal plant *Podophyllum peltatum*. *Evolution* 1989;**43**:540−7.
3. Salvaudon L, Giraud T, Shykoff JA. Genetic diversity in natural populations: a fundamental component of plant-microbe interactions. *Curr Opin Plant Biol* 2008;**11**:135−43.
4. Buckling A, Rainey PB. The role of parasites in sympatric and allopatric host diversification. *Nature* 2002;**420**:496−9.
5. Thompson JN. *The geographic mosaic of coevolution*. Chicago: University of Chicago Press; 2005.

Downloaded From Vet-eBooks.com

6. Laine AL. Role of coevolution in generating biological diversity: spatially divergent selection trajectories. *J Exp Bot* 2009;**60**:2957−70.
7. Paterson S, Vogwill T, Buckling A, et al. Antagonistic coevolution accelerates molecular evolution. *Nature* 2010;**464**:275−8.
8. Janzen DH. When is it coevolution? *Evolution* 1978;**34**:611−2.
9. Hudson P, Dobson A, Lafferty K. Is a healthy ecosystem one that is rich in parasites? *Trends Ecol Evol* 2006;**21**:381−5.
10. Lively CM. Migration, virulence, and the geographic mosaic of adaptation by parasites. *Am Nat* 1999;**153**:S34−47.
11. Jaenike J. An hypothesis to account for the maintenance of sex within populations. *Evol Theory* 1978;**3**:191−4.
12. Prenter J, MacNeil C, Dick JTA, Dunn AM. Roles of parasites in animal invasions. *Trends Ecol Evol* 2004;**19**:385−90.
13. Buckling A, Rainey PB. Antagonistic coevolution between a bacterium and a bacteriophage. *Proc R Soc Lond Ser B Biol Sci* 2002;**269**:931−6.
14. Forde SE, Thompson JN, Bohannan BJM. Adaptation varies through space and time in a coevolving host- parasitoid interaction. *Nature* 2004;**431**:841−4.
15. Morgan AD, Gandon S, Buckling A. The effect of migration on local adaptation in a coevolving host-parasite system. *Nature* 2005;**437**:253−6.
16. Koskella B, Lively CM. Advice of the rose: experimental coevolution of a trematode parasite and its snail host. *Evolution* 2007;**61**:152−9.
17. Jokela J, Dybdahl MF, Lively CM. The maintenance of sex, clonal dynamics, and host-parasite coevolution in a mixed population of sexual and asexual snails. *Am Nat* 2009; **174**:S43−53.
18. Gandon S, Van Zandt PA. Local adaptation and host-parasite interactions. *Trends Ecol Evol* 1998;**13**:214−6.
19. Nuismer SL, Thompson JN, Gomulkiewicz R. Coevolution between hosts and parasites with partially overlapping geographic ranges. *J Evol Biol* 2003;**16**:1337−45.
20. Nuismer SL, Goodnight C. Parasite local adaptation in a geographic mosaic. *Evolution* 2009;**60**:24−30.
21. Hamilton WD. Pathogens as causes of genetic diversity in their host populations. In: Anderson RM, May RM, editors. *Population biology of infectious diseases*. New York: Springer-Verlag; 1982.
22. Nee S. Antagonistic co-evolution and the evolution of genotypic randomization. *J Theor Biol* 1989;**140**:499−518.
23. Kniskern JM, Barrett LG, Bergelson J. Maladaptation in wild populations of the generalist plant pathogen *Pseudomonas syringae*. *Evolution* 2011;**65**:818−30.
24. Thompson JN, Burdon JJ. Gene-for-gene coevolution between plants and parasites. *Nature* 1992;**360**:121−5.
25. Parker MA. Pathogens and sex in plants. *Evol Ecol* 1994;**8**:560−84.
26. Sasaki A. Host-parasite coevolution in a multilocus gene-for-gene system. *Proc R Soc Lond Ser B Biol Sci* 2000;**267**:2183−8.
27. Boots M, Begon M. Trade-offs with resistance to a granulosis-virus in the Indian Meal Moth, examined by a laboratory evolution experiment. *Funct Ecol* 1993;**7**:528−34.
28. Ferdig MT, Beerntsen BT, Spray FJ, Li JY, Christensen BM. Reproductive costs associated with resistance in a mosquito-filarial worm system. *Am J Trop Med* 1993;**49**:756−62.
29. Fellowes MDE, Kraaijeveld AR, Godfray HCJ. Trade-off associated with selection for increased ability to resist parasitoid attack in *Drosophila melanogaster*. *Proc R Soc B Biol Sci* 1998;**265**:1553−8.

Downloaded From Vet-eBooks.com

30. Langand J, Jourdane J, Coustau C, Delay B, Morand S. Cost of resistance, expressed as a delayed maturity, detected in the host−parasite system *Biomphalaria glabrata Echinostoma caproni*. *Heredity* 1998;**80**:320−5.
31. Bahri B, Kaltz O, Leconte M, de Vallavieille-Pope C, Enjalbert J. Tracking costs of virulence in natural populations of the wheat pathogen, *Puccinia striiformis* f.sp.*tritici*. *BMC Evol Biol* 2009;**9**.
32. Grim T, Rutila J, Cassey P, Hauber ME. The cost of virulence: an experimental study of egg eviction by brood parasitic chicks. *Behav Ecol* 2009;**20**:1138−46.
33. Huang YJ, Balesdent MH, Li ZQ, Evans N, Rouxel T, Fitt BDL. Fitness cost of virulence differs between the AvrLm1 and AvrLm4 loci in *Leptosphaeria maculans* (phoma stem canker of oilseed rape). *Eur J Plant Pathol* 2010;**126**:279−91.
34. Hamilton WD. Sex vs. non-sex vs. parasite. *Oikos* 1980;**35**:282−90.
35. Frank SA. Recognition and polymorphism in host-parasite genetics. *Philos Trans R Soc Lond B* 1994;**346**:283−93.
36. Bell G. *The masterpiece of nature: the evolution and genetics of sexuality*. Berkeley: University of California Press; 1982.
37. Carroll L. *Through the looking-glass*. London: Macmillan; 1871.
38. Woolhouse MEJ, Webster JP. In search of the Red Queen. *Parasitol Today* 2000;**16**: 506−8.
39. Seger J. Dynamics of some simple host-parasite models with more than two genotypes in each species. *Philos Trans R Soc Lond Ser B Biol Sci* 1988;**319**:541−55.
40. Morand S, Manning SD, Woolhouse MEJ. Parasite-host coevolution and geographic patterns of parasite infectivity and host susceptibility. *Proc R Soc Lond Ser B Biol Sci* 1996;**263**:119−28.
41. Price MV, Waser NM. Population structure, frequency-dependent selection, and the maintenance of sexual reproduction. *Evolution* 1982;**36**:35−43.
42. Hamilton WD, Axelrod R, Tanese R. Sexual reproduction as adaptation to resist parasites (a review). *Proc Natl Acad Sci USA* 1990;**87**:3566−73.
43. Peters AD, Lively CM. The Red Queen and fluctuating epistasis: a population genetic analysis of antagonistic coevolution. *Am Nat* 1999;**154**:393−405.
44. Grosberg RK, Hart MW. Mate selection and the evolution of highly polymorphic self/nonself recognition genes. *Science* 2000;**289**:2111−4.
45. Luijckx P, Fienberg H, Duneau D, Ebert D. A matching-allele model explains host resistance to parasites. *Curr Biol* 2013;**23**:1085−8.
46. Flor HH. The complementary genetic systems in flax and flax rust. *Adv Genet* 1956;**8**: 29−54.
47. Agrawal A, Lively CM. Infection genetics: gene-for-gene versus matching-alleles models and all points in between. *Evol Ecol Res* 2002;**4**:79−90.
48. Agrawal AF, Lively CM. Modelling infection as a two-step process combining gene-for-gene and matching-allele genetics. *Proc R Soc Lond Ser B Biol Sci* 2003;**270**:323−34.
49. Salathe M, Kouyos RD, Regoes RR, Bonhoeffer S. Rapid parasite adaptation drives selection for high recombination rates. *Evolution* 2008;**62**:295−300.
50. Henter HJ, Via S. The potential for coevolution in a host-parasitoid system. I. Genetic variation within an aphid population in susceptibility to a parasitic wasp. *Evolution* 1995; **49**:257−65.
51. Kraaijeveld AR, Van Alphen JJM, Godfray HCJ. The coevolution of host resistance and parasitoid virulence. *Parasitology* 1998;**116**:S29−45.
52. Webster JP, Woolhouse MEJ. Cost of resistance to *Schistosoma mansoni* in the snail intermediate host *Biomphalaria glabrata*. *Trans R Soc Trop Med Hyg* 1998;**92**:367.

Downloaded From Vet-eBooks.com

53. Little TJ, Ebert D. Associations between parasitism and host genotype in natural populations of *Daphnia* (Crustacea : Cladocera). *J Anim Ecol* 1999;**68**:134−49.
54. Salvaudon L, Heraudet V, Shykoff JA. Parasite-host fitness trade-offs change with parasite identity: genotype-specific interactions in a plant-pathogen system. *Evolution* 2005;**59**: 2518−24.
55. Blanford S, Thomas MB, Pugh C, Pell JK. Temperature checks the Red Queen? Resistance and virulence in a fluctuating environment. *Ecol Lett* 2003;**6**:2−5.
56. Carius HJ, Little TJ, Ebert D. Genetic variation in a host-parasite association: potential for coevolution and frequency-dependent selection. *Evolution* 2001;**55**:1136−45.
57. Schulenburg H, Ewbank J. Diversity and specificity in the interaction between *Caenorhabditis elegans* and the pathogen *Serratia marcescens*. *BMC Evol Biol* 2004;**4**:49.
58. Lambrechts L, Halbert J, Durand P, Gouagna L, Koella J. Host genotype by parasite genotype interactions underlying the resistance of anopheline mosquitoes to *Plasmodium falciparum*. *Malar J* 2005;**4**:3.
59. Rauch G, Kalbe M, Reusch TBH. One day is enough: rapid and specific host−parasite interactions in a stickleback-trematode system. *Biol Lett* 2006;**2**:382−4.
60. Yourth CP, Schmid-Hempel P. Serial passage of the parasite *Crithidia bombi* within a colony of its host, *Bombus terrestris*, reduces success in unrelated hosts. *Proc R Soc B Biol Sci* 2006;**273**:655−9.
61. Duffy S, Burch CL, Turner PE. Evolution of host specificity drives reproductive isolation among RNA viruses. *Evolution* 2007;**61**:2614−22.
62. Nidelet T, Kaltz O. Direct and correlated responses to selection in a host-parasite system: testing for the emergence of genotype specificity. *Evolution* 2007;**61**:1803−11.
63. Anderson RM, May RM. Coevolution of hosts and parasites. *Parasitology* 1982;**85**: 411−26.
64. Frank SA. Models of parasite virulence. *Q Rev Biol* 1996;**71**:37−78.
65. Day T. Parasite transmission modes and the evolution of virulence. *Evolution* 2001;**55**: 2389−400.
66. Bolker BM, Nanda A, Shah D. Transient virulence of emerging pathogens. *J R Soc Interface* 2010;**7**:811−22.
67. Boots M, Hudson PJ, Sasaki A. Large shifts in pathogen virulence relate to host population structure. *Science* 2004;**303**:842−4.
68. Bremermann HJ, Pickering J. A game-theoretical model of parasite virulence. *J Theor Biol* 1983;**100**:411−26.
69. Brown SP, Hochberg ME, Grenfell BT. Does multiple infection select for raised virulence? *Trends Microbiol* 2002;**10**:401−5.
70. Bérénos C, Schmid-Hempel P, Wegner KM. Evolution of host resistance and trade-offs between virulence and transmission potential in an obligately killing parasite. *J Evol Biol* 2009;**22**:2049−56.
71. Zund P, Lebek G. Generation time-prolonging R plasmids: correlation between increases in the generation time of *Escherichia coli* caused by R plasmids and their molecular size. *Plasmid* 1980;**3**:65−9.
72. Bouma JE, Lenski RE. Evolution of a bacteria/plasmid association. *Nature* 1988;**335**: 351−2.
73. Modi RI, Adams J. Coevolution in bacterial-plasmid populations. *Evolution* 1991;**45**: 656−67.
74. Nowak MA, May RM. Superinfection and the evolution of parasite virulence. *Proc R Soc Lond Ser B Biol Sci* 1994;**255**:81−9.
75. Turner PE, Chao L. Prisoner's dilema in an RNA virus. *Nature* 1999;**398**:441−3.

Downloaded From Vet-eBooks.com

76. Griffin AS, West SA, Buckling A. Cooperation and competition in pathogenic bacteria. *Nature* 2004;**430**:1024−7.
77. Harrison F, Browning LE, Vos M, Buckling A. Cooperation and virulence in acute *Pseudomonas aeruginosa* infections. *BMC Biol* 2006;**4**.
78. Frank SA. The origin of synergistic symbiosis. *J Theor Biol* 1995;**176**:403−10.
79. Dale C, Plague GR, Wang B, Ochman H, Moran NA. Type III secretion systems and the evolution of mutualistic endosymbiosis. *Proc Natl Acad Sci USA* 2002;**99**: 12397−402.
80. Buttner D, Bonas U. Port of entry − the type III secretion translocon. *Trends Microbiol* 2002;**10**:186−92.
81. Galan JE, Collmer A. Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* 1999;**284**:1322−8.
82. Brown MJF, Loosli R. Schmid-Hempel P and Condition-dependent expression of virulence in a trypanosome infecting bumblebees. *Oikos* 2000;**91**:421−7.
83. Pulkkinen K, Ebert D. Host starvation decreases parasite load and mean host size in experimental populations. *Ecology* 2004;**85**:823−33.
84. Bedhomme S, Agnew P, Vital Y, Sidobre C, Michalakis Y. Prevalence-dependent costs of parasite virulence. *PLoS Biol* 2005;**3**:1403−8.
85. Lively CM. The ecology of virulence. *Ecol Lett* 2006;**9**:1089−95.
86. De Roode JC, Pedersen AB, Hunter MD, Altizer S. *Host plant species affects virulence in monarch butterfly parasites*. Oxford, UK: Blackwell; 2008.
87. Gandon S, Michalakis Y. Local adaptation, evolutionary potential and host-parasite coevolution: interactions between migration, mutation, population size and generation time. *J Evol Biol* 2002;**15**:451−62.
88. Pieterse CJ, de Jonge R, Berendsen RL. *Soil Borne Supremacy* 2016;**21**:171−3.
89. Gall CA, Reif KE, Scoles GA, et al. The bacterial microbiome of *Dermacentor andersoni* ticks influences pathogen susceptibility. *ISME J* 2016;**10**:1846−55.
90. Ritpitakphong U, Falquet L, Vimoltust A, Berger A, Metraux J-P, L'Haridon F. The microbiome of the leaf surface of Arabidopsis protects against a fungal pathogen. *New Phytol* 2016;**2010**:1033−43.
91. Brockhurst MA, Koskella B. Experimental coevolution of species interactions. *Trends Ecol Evol* 2013;**28**:367−75.
92. Price PW. *Evolutionary biology of parasites*. Princeton, NJ: Princeton University Press; 1980.
93. Ebert D. Virulence and local adaptation of a horizontally transmitted parasite. *Science* 1994;**265**:1084−6.
94. Thompson JN. Specific hypotheses on the geographic mosaic of coevolution. *Am Nat* 1999;**153**:S1−14.
95. Gomulkiewicz R, Thompson JN, Holt RD, Nuismer SL, Hochberg ME. Hot spots, cold spots, and the geographic mosaic theory of coevolution. *Am Nat* 2000;**156**:156−74.
96. Gandon S. Local adaptation and the geometry of host-parasite coevolution. *Ecol Lett* 2002; **5**:246−56.
97. Greischar MA, Koskella B. A synthesis of experimental work on parasite local adaptation. *Ecol Lett* 2007;**10**(5):418−34.
98. Thompson JN. *The coevolutionary process*. University of Chicago Press; 1994.
99. Cory JS, Myers JH. The ecology and evolution of insect baculoviruses. *Annu Rev Ecol Evol Syst* 2003;**34**:239−72.
100. Cory JS, Myers JH. Adaptation in an insect host-plant pathogen interaction. *Ecol Lett* 2004;**7**:632−9.

Downloaded From Vet-eBooks.com

101. Koskella B, Thompson JN, Preston GM, Buckling A. Local biotic environment shapes the spatial scale of bacteriophage adaptation to bacteria. *Am Nat* 2011;**177**:440–51.
102. King KC, Delph LF, Jokela J, Lively CM. The geographic mosaic of sex and the Red Queen. *Curr Biol* 2009;**19**:1438–41.
103. Abt MC, Pamer EG. Commensal bacteria mediated defenses against pathogens. *Curr Opin Immunol* 2014;**29**:16–22.
104. Clay K. Defensive symbiosis: a microbial perspective. *Funct Ecol* 2014;**28**:293–8.
105. Kemen E. Microbe-microbe interactions determine oomycete and fungal host colonization. *Curr Opin Plant Biol* 2014;**20**:75–81.
106. Round JL, Mazmanian SK. The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* 2009;**9**:313–23.
107. Koch H, Schmid-Hempel P. Socially transmitted gut microbiota protect bumble bees against an intestinal parasite. *Proc Natl Acad Sci USA* 2011;**108**:19288–92.
108. Koch H, Schmid-Hempel P. Gut microbiota instead of host genotype drive the specificity in the interaction of a natural host-parasite system. *Ecol Lett* 2012;**15**:1095–103.
109. Rovenich H, Boshoven JC, Thomma BPHJ. Filamentous pathogen effector functions: of pathogens, hosts and microbiomes. *Curr Opin Plant Biol* 2014;**20**:96–103.
110. Schreiber F, Arasteh JM, Lawley TD. Pathogen resistance mediated by IL-22 signaling at the epithelial–microbiota interface. *J Mol Biol* 2015;**427**:3676–82.
111. Lopez-Pascua LDC, Buckling A. Increasing productivity accelerates host-parasite coevolution. *J Evol Biol* 2008;**21**:853–60.
112. Lopez Pascua L, Hall AR, Best A, Morgan AD, Boots M, Buckling A. Higher resources decrease fluctuating selection during host-parasite coevolution. *Ecol Lett* 2014;**17**: 1380–8.
113. Forde SE, Thompson JN, Bohannan BJM. Gene flow reverses an adaptive cline in a coevolving host-parasitoid interaction. *Am Nat* 2007;**169**:794–801.
114. Lopez Pascua L, Gandon S, Buckling A. Abiotic heterogeneity drives parasite local adaptation in coevolving bacteria and phages. *J Evol Biol* 2012;**25**:187–95.
115. Gomez P, Buckling A. Bacteria-phage antagonistic coevolution in soil. *Science* 2011;**332**: 106–9.
116. Zhang Q-G, Buckling A. Antagonistic coevolution limits population persistence of a virus in a thermally deteriorating environment. *Ecol Lett* 2011;**14**:282–8.
117. Gaba S, Ebert D. Time-shift experiments as a tool to study antagonistic coevolution. *Trends Ecol Evol* 2009;**24**:226–32.
118. Elena SF, Lenski RE. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet* 2003;**4**:457–69.
119. Buckling A, Maclean RC, Brockhurst MA, Colegrave N. The Beagle in a bottle. *Nature* 2009;**457**:824–9.
120. Bull JJ, Mollineux IJ, Rice WR. Selection of benevolence in a host-parasite system. *Evolution* 1991;**45**:875–82.
121. Morgan AD, Buckling A. Relative number of generations of hosts and parasites does not influence parasite local adaptation in coevolving populations of bacteria and phages. *J Evol Biol* 2006;**19**:1956–63.
122. Brockhurst MA, Morgan AD, Rainey PB, Buckling A. Population mixing accelerates coevolution. *Ecol Lett* 2003;**6**:975–9.
123. Morgan AD, Bonsall MB, Buckling A. Impact of bacterial mutation rates in coevolutionary dynamics between bacteria and phage. *Evolution* 2010;**10**:2980–7.
124. Decaestecker E, Gaba S, Raeymaekers JAM, et al. Host-parasite 'Red Queen' dynamics archived in pond sediment. *Nature* 2007;**450**:870–3.

Downloaded From Vet-eBooks.com

125. Koskella B. Phage-mediated selection on microbiota of a long-lived host. *Curr Biol* 2013; **23**:1256−60.
126. Koskella B. Bacteria-phage interactions across time and space: merging local adaptation and time-shift experiments to understand phage evolution*. *Am Nat* 2014;**184**:S9−21.
127. Hutson V, Law R. Evolution of recombination in populations experiencing frequency-dependent selection with time delay. *Proc R Soc B Biol Sci* 1981;**213**:345−59.
128. Clarke B. The ecological relationships of host-parasite relationships. In: Taylor AER, Muller R, editors. *Genetic aspects of host-parasite relationships*. Oxford: Blackwell; 1976. p. 87−103.
129. Bremermann HJ, Fiedler B. On the stability of polymorphic host-pathogen populations. *J Theor Biol* 1985;**117**:621−31.
130. Hamilton WD. Haploid dynamic polymorphism in a host with matching parasites: effects of mutation/subdivision, linkage, and patterns of selection. *J Hered* 1993;**84**:328−38.
131. Dybdahl MF, Lively CM. Host-parasite coevolution: evidence for rare advantage and time-lagged selection in a natural population. *Evolution* 1998;**52**:1057−66.
132. Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J. Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature* 1999;**400**:667−71.
133. Wolinska J, Spaak P. The cost of being common: evidence from natural *Daphnia* populations. *Evolution* 2009;**63**:1893−901.
134. Koskella B, Lively CM. Evidence for negative frequency-dependent selection during experimental coevolution of a freshwater snail and a sterilizing trematode. *Evolution* 2009; **63**:2213−21.
135. Lively CM. Adaptation by a parasitic trematode to local populations of its snail host. *Evolution* 1989;**43**:1663−71.
136. Dybdahl MF, Lively CM. Host-parasite interactions:infection of common clones in natural populations of a freshwater snail. *Proc R Soc Lond Ser B Biol Sci* 1995;**260**:99−103.
137. Capaul M, Ebert D. Parasite-mediated selection in experimental *Daphnia magna* populations. *Evolution* 2003;**57**:249−60.
138. Duncan AB, Little TJ. Parasite-driven genetic change in a natural popualtion of *Daphnia*. *Evolution* 2007;**61**:796−803.
139. Duffy MA, Brassil CE, Hall SRT AJ, Caceres CE, Conner JK. Parasite-mediated disruptive selection in a natural *Daphnia* population. *BMC Evol Biol* 2008;**8**:80.
140. Ebert D. Host-parasite coevolution: insights from the *Daphnia*-parasite model system. *Curr Opin Microbiol* 2008;**11**:290−301.
141. Kawecki TJ, Ebert D. Conceptual issues in local adaptation. *Ecol Lett* 2004;**7**:1225−41.
142. Parker MA. Local population differentiation for compatibility in an annual legume and its host-specific fungal pathogen. *Evolution* 1985;**39**:713−23.
143. Roy BA. Differentiating the effects of origin and frequency in reciprocal transplant experiments used to test negative frequency-dependent selection hypothesis. *Oecologia* 1998;**115**:73−83.
144. Kaltz O, Shykoff JA. Local adaptation in host-parasite systems. *Heredity* 1998;**81**: 361−70.
145. Hoeksema JD, Forde SE. A meta-analysis of factors affecting local adaptation between interacting species. *Am Nat* 2008;**171**:275−90.
146. Gandon S, Capowiez Y, Dubois Y, Michalakis Y, Olivieri I. Local adaptation and gene-for-gene coevolution in a metapopulation model. *Proc R Soc Lond Ser B Biol Sci* 1996; **263**:1003−9.
147. Weiblen GD, Bush GL. Speciation in fig pollinators and parasites. *Mol Ecol* 2002;**11**: 1573−8.

Downloaded From Vet-eBooks.com

148. Zietara MS, Lumme J. Speciation by host switch and adaptive radiation in a fish parasite genus *Gyrodactylus* (*Monogenea, Gyrodactylidae*). *Evolution* 2002;**56**:2445−58.
149. Sorenson MD, Sefc KM, Payne RB. Speciation by host switch in brood parasitic indigobirds. *Nature* 2003;**424**:928−31.
150. Hafner MS, Page RDM. Molecular phylogenies and host-parasite cospeciation - gophers and lice as a model. *Philos Trans R Soc Lond Ser B Biol Sci* 1995;**349**:77−83.
151. Storfer A, Alfaro ME, Ridenhour BJ, et al. Phylogenetic concordance analysis shows an emerging pathogen is novel and endemic. *Ecol Lett* 2007;**10**:1075−83.
152. Shafer ABA, Williams GR, Shutler D, Rogers REL, Stewart DT. Cophylogeny of Nosema (*Microsporidia: Nosematidae*) and bees (*Hymenoptera: Apidae*) suggests both cospeciation and a host-switch. *J Parasitol* 2009;**95**:198−203.
153. Hafner MS, Nadler SA. Phylogenetic trees support the coevolution of parasites and their hosts. *Nature* 1988;**332**:258−9.
154. Maynard Smith J. *The evolution of sex*. Cambridge: Cambridge University Press; 1978.
155. Hurst LD, Peck JR. Recent advances in understanding of the evolution and maintenance of sex. *Trends Ecol Evol* 1996;**11**:A46−52.
156. Bell G. Recombination and the immortality of the germ line. *J Evol Biol* 1988;**1**:67−82.
157. Maynard Smith J. The maintenance of sex. In: Ridley M, editor. *Evolution*. Oxford: Oxford University Press; 1997.
158. Kondrashov A. Deleterious mutations and the evolution of sexula reproduction. *Nature* 1988;**336**:435−40.
159. de Visser JAGM, Hoekstra RF, van den Ende H. The effect of sex and deleterious mutations on fitness in *Chlamydomonas*. *Proc R Soc B Biol Sci* 1996;**263**:193−200.
160. Lively CM, Craddock C, Vrijenhoek RC. Red Queen hypothesis supported by parasitism in sexual and clonal fish. *Nature* 1990;**344**:864−6.
161. Fischer O, Schmid-Hempel P. Selection by parasites may increase host recombination frequency. *Biol Lett* 2005;**1**:193−5.
162. Giraud A, Matic I, Tenaillon O, et al. Costs and benefits of high mutatation rates: adaptive evolution of bacteria in the mouse gut. *Science* 2001;**291**:2606−8.
163. Pal C, Macia MD, Oliver A, Schachar I, Buckling A. Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* 2007;**450**:1079−81.
164. Ebert D, Bull JJ. Challenging the trade-off model for the evolution of virulence: is virulence management feasible? *Trends Microbiol* 2003;**11**:15−20.
165. Wilson DS, Yoshimura J. On the coexistence of specialists and generalists. *Am Nat* 1994; **144**:692.
166. Regoes RR, Nowak MA, Bonhoeffer S. Evolution of virulence in a heterogeneous host population. *Evolution* 2000;**54**:64−71.
167. Morgan AD, Brockhurst MA, Lopez-Pascua LDC, Pal C, Buckling A. Differential impact of simultaneous migration on coevolving hosts and parasites. *BMC Evol Biol* 2007;**7**.
168. Tompkins DM, White AR, Boots M. Ecological replacement of native red squirrels by invasive greys driven by disease. *Ecol Lett* 2003;**6**:189−96.
169. Torchin ME, Mitchell CE. Parasites, pathogens, and invasions by plants and animals. *Front Ecol Environ* 2004;**2**:183−90.
170. Antonovics J, Hood M, Partain J. The ecology and genetics of a host shift: microbotryum as a model system. *Am Nat* 2002;**160**:S40−53.
171. López-Villavicencio M, Enjalbert J, Hood ME, Shykoff JA, Raquin C, Giraud T. The another smut disease on *Gypsophila repens*: a case of parasite sub-optimal performance following a recent host shift? *J Evol Biol* 2005;**18**:1293−303.

Downloaded From Vet-eBooks.com

172. Wright A, Hawkins CH, Anggard EE, Harper DR. A controlled clinical trial of a therapeutic bacteriophage preparation in chronic otitis due to antibiotic-resistant *Pseudomonas aeruginosa*; a preliminary report of efficacy. *Clin Otolaryngol* 2009;**34**:349−57.
173. Levin BR, Bull JJ. Population and evolutionary dynamics of phage therapy. *Nat Rev Microbiol* 2004;**2**:166−73.
174. Williams PD. Darwinian interventions: taming pathogens through evolutionary ecology. *Trends Parasitol* 2010;**26**:83−92.
175. Nesse RM, Bergstrom CT, Ellison PT, et al. Making evolutionary biology a basic science for medicine. *Proc Natl Acad Sci USA* 2010;**107**:1800−7.
176. Dobzhansky T. Nothing in biology makes sense except in the light of evolution. *Am Biol Teach* 1973;**35**:125−9.

Downloaded From Vet-eBooks.com

Microbes as Tracers of Past Human Demography and Migrations

7

J.-P. Rasigade[1,2], A. Gilabert[3], T. Wirth[1,4]
[1]Ecole Pratique des Hautes Etudes, Paris Sciences et Lettres, Paris, France; [2]CIRI, International Center for Infectiology Research, INSERM U1111-CNRS UMR5308, ENS Lyon-Université Lyon 1, Hospices Civils de Lyon, Lyon, France; [3]Laboratoire MIVEGEC, UMR 5290, IRD-CNRS-UM, Montpellier, France; [4]Institut de Systématique, Evolution, Biodiversité, UMR-CNRS 7205, Muséum National d'Histoire Naturelle, Université Pierre et Marie Curie, Ecole Pratique des Hautes Etudes, Sorbonne Universités, Paris, France

1. Introduction

A major aspect of human evolutionary biology consists in disentangling human origins and migrations. To address this aim, human population genetics has been directly investigated using polymorphic markers such as proteins, mtDNA, Y-chromosome, microsatellites, or single nucleotide polymorphisms (SNPs) (see Refs. 1–3 for reviews). These studies, combined to morphological, anthropological, and linguistic ones have led to the formulation of a standard model of modern human evolution. This theory advocates that humans originated in East Africa and dispersed, first throughout much of Africa about 100,000–150,000 years ago,[4,5] and subsequently—between 70,000 and 40,000 years ago—into Asia and then Europe.[2,6] The settlements of the Americas and Oceania occurred later through several migrations out of Asia.[7–9] These successive waves of migrations resulted in a relatively low genetic diversity within modern human populations and in a decrease of genetic diversity from the horn of Africa. The genetic differentiation increases with geographical distances following an isolation-by-distance (IBD) model, but remains low ($F_{st} < 2\%$). Therefore, human genetic studies are often weakly resolved and moderately informative.[10–12]

As stated in the preceding paragraph, the use of human genetic markers has contributed to the understanding of human evolution but has also failed to elucidate some recent features. Several issues remained controversial for a long time, such as the timing, the source, and number of migrations to America[13] and to Oceania.[14] In addition, relationships between closely related populations are difficult to decipher because of their too recent divergence. Indeed, direct inference of human evolutionary history is limited because of the low genetic variability due to the strong genetic bottlenecks that humans were subjected to during migrations. Other techniques such as microsatellites or SNPs, supposedly more variable than other markers, also present technical

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00007-X
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

limitations to resolve human migrations.[15] However, nowadays whole-genome sequencing of present-day humans or ancient DNA sequencing can dramatically increase the resolution and improve our understanding of human evolution.

To overcome these hindrances, an alternative is to focus on human pathogens since they have coevolved with humans and reflect their evolutionary histories.[16] Pathogens present generally higher mutation rates and shorter generation times than humans.[17] Thus, their populations are more diversified genetically, making the study of their population structure more informative than that of human. However, not all microbes are good candidates to infer host evolution and their efficacy depends on several factors.[17] In addition, the choice of the pathogen will also be influenced by the time scale of the study.[18]

Several pathogens have proven their usefulness in deciphering human migrations and origins.[16] In particular, their study allows the distinction between closely related groups of humans, which previously was not directly possible due to a lack of resolution of markers and/or a sampling failure for instance.[14,15] In this review, after a brief section on some advantages and disadvantages of pathogens in the context of human host−history inference, we illustrate pathogens' utility with some relevant examples that pointed out congruence or discrepancies with human migratory history.

2. Using Pathogens as Genetic Tracers for Host History

Parasites have often been used to infer their host evolutionary history,[17] usually using phylogeographical analyses, due to their narrow relationships with their hosts as well as their generally higher levels of genetic diversity. However, even if both protagonists share a common history, their genealogies are not necessarily similar[18,19] and several evolutionary mechanisms can lead to identical gene trees.[20] Therefore, microbes have to be carefully chosen to be relevant in this context, some pathogen traits being of particular importance to correctly infer host history. Crucial features and parameters that determine their usefulness are degree of intimacy with their host species and mode of transmission, as well as mutation and recombination rates.[17,18,20]

First of all, parasites without any secondary hosts or free-living stages are preferable over those with such complex life cycle, which may evolve independently from the host of interest.[18] Also, pathogens are more relevant to infer host history if they are persistent and transmitted vertically (from parents to children). When pathogens are transmitted through an epidemic, their population structure tends to reflect their own demographic history (frequent bottlenecks followed by rapid expansions) than those of their host.[19] However, low rates of horizontal transmission, if occurring within and not among divergent populations, in parallel with vertical transmission, will not lead to incongruence between parasite and host trees.[16,18] Therefore, in the selection of a pathogen species as an inferential tool, one has to first know its life cycle and its mode of transmission, which can be achieved by means of ecological surveys, experimentations, and within-family studies.[21]

Downloaded From Vet-eBooks.com

The rate of molecular evolution also greatly determines the efficacy of microbes as tracers. If the mutation rate is too low, the resolution of phylogenies can be crude and recent events may not be detected due to a lack of signal. Thus, in such cases, the use of parasites to infer host genealogies is not obvious. The opposite is also questionable: with a mutation rate that is too high, one can overlook information due to saturation at informative sites and homoplasy. In addition, depending on the mutation rate of their DNA, studies of pathogens give insight into past or recent events in their host histories. A wrong estimation of the mutation rate may lead to misinterpretations.

Another important parameter is the recombination rate. Indeed, recombination, even at a low rate, leads to intermediate genotypes, larger terminal branches, and an underestimation of the time to the most recent common ancestor (MRCA).[22] However, recombination generally occurs between related populations, which permits the maintenance of the genetic structure even if the signal is weak. Several methods have been developed to estimate the recombination rate (see, e.g., Refs. 23–25) including coalescent and Monte Carlo–based simulation methods.[26–29] Otherwise, the homoplasy test[30] or the compatibility matrices test[31] detect and estimate the frequency of recombination events (see, e.g., Ref. 32). When evidence of recombination is found, other approaches such as a Bayesian clustering method, which can deal with recombination, is preferable.

Finally, selection can also dramatically reduce the reliability of molecular phylogenies since populations can be clustered together because they are under identical selection regimes despite their distinct evolutionary histories (Ref. 16). Another illustration can be found in a study by Devi et al.[33] in which *H. pylori* population structure has been investigated from housekeeping genes and the *cag* pathogenecity island (*cag*PAI), which is under selective pressure. In this study, all Peruvian strains harbored a "western"-type *cag*PAI, suggesting a European origin, while the analysis based on the housekeeping genes revealed that some clustered with the hpAmerind population (see the following) suggesting an Asian origin for these strains. Hence, an analysis based on the sole *cag*PAI would not include all information. Selection can be detected in protein-coding genes by comparing the number of nonsynonymous amino acid changes (d_N) with the number of synonymous amino acid changes (d_S) with standard[34] (DNAsp) or more sophisticated[35] (PAML) tools. If the ratio d_N/d_S is equal to 1, the gene is under neutral evolution while if it is different from 1, the gene is either under purifying (<1) or directional (>1) selection. Methods are developed to detect loci under selection (e.g., BayeScan[36] or PCadapt[37] in intra-specific data sets) and could be used when whole-genome sequences of parasites are available.

To overcome some of these discrepancies (such as mutation and recombination rates or selection), reconstructing phylogenies from several genes is preferable. Incongruence among individual gene signals can be tested, for example, with the partition homogeneity test.[38] When this test reveals homogeneity among datasets, the different genes can be concatenated. In addition, to confirm pathogen efficiency to infer host evolutionary history, it would be advisable to directly compare host and pathogen phylogenies by collecting data from the same material.

Downloaded From Vet-eBooks.com

3. Candidates

3.1 Bacteria

3.1.1 Helicobacter pylori

The ubiquitous bacterium *H. pylori* has been shown to be a powerful tracer of human population structure[15,39] and it is one of the most studied pathogens in human history inferences. *Helicobactor pylori* is a Gram-negative bacteria that infects human stomachs and is associated with gastrointestinal diseases such as gastritis, ulcers, or cancers although infections are relatively infrequent and mostly benign.[40] Prevalence of the infection exceeds 50% of the human population but decreases in industrialized countries.[21,40] For a long time, *H. pylori* was thought to be mainly transmitted vertically during childhood.[41–43] However, investigations on the transmission of this bacterium in both developed and developing countries revealed that horizontal transmission might not be negligible and might depend on socioeconomic status.[21,44] Mixed infections of *H. pylori* are not rare but generally involve one dominant strain.[21,44,45] This bacterium species shows an unusually high level of genetic diversity which may result from a combination of high mutation rates, frequent recombination events, and a continuous acquisition of new strains.[32,44,46,47] The genome-wide mutation rate of *H. pylori* has been estimated using pairs of sequential sampling to be around $0.5–6.5 \times 10^{-5}$ per year per site, and the recombination rate is estimated to be around 5.5×10^{-5} recombination events per initiation site and year.[48] *Helicobactor pylori* appeared to be nearly panmictic[43] although several studies revealed phylogeographical differences. This apparent contradiction could be explained by frequent recombinations between geographically related strains so that the population structure can be maintained.[44]

The suggestion that *H. pylori* has coevolved with humans and that its population structure reflects human migrations was first reported in studies that investigated the genetic differences between bacterial populations from distinct areas.[32,49] Since then, several studies allowed a fine timing of the relationships between *H. pylori* and humans and the elucidation of ancient human migrations. Linz et al.[50] documented a linear relationship between geographical distance from Africa and the microbial genetic diversity. IBD patterns were also observed both at a global and a local (European) scale. Similar correlations were obtained in humans, highlighting an intimate association between *H. pylori* and humans over a long period of time. Ramachandran et al.[12] observed, in addition to such an IBD pattern, a decrease in the expected heterozygosity (estimated from 783 microsatellite loci) with distance from Addis Ababa. Using simulations, the authors evidenced that this pattern can be explained by serial founder effects starting at a single origin thus confirming sequential waves of migration during modern human history. Linz et al.[50] also investigated the origin and demography of *H. pylori* by means of demographically explicit genetic simulations. Three alternative scenarios were tested for the origin of *H. pylori*: an East African, a South African, and an East Asian origin. The best model (based on the proportion of total variance explained by the model and the Akaike information criterion) argued in favor of an

Downloaded From Vet-eBooks.com

east African origin. The simulations also indicated that *H. pylori* spread from Africa about 58,000 years ago, which is consistent with the dating of human migrations out of Africa.[10] Linz et al.[50] concluded that all these parallels between bacteria and humans, observed at both global and local scales, reflect an expansion of *H. pylori* via ancient human migrations and genetic admixture after horizontal transmission or through recent migrations. Moreover, this diversification of *H. pylori*, concomitant with out-of-Africa migrations, suggested that humans were already infected before their initial migrations.[50] This hypothesis received further support as coalescent analyses of African *H. pylori* showed that lineages infecting southern African San ethnic groups diverged from other lineages 88,000 to 116,000 years before present, indicating that *H. pylori* emergence largely predated out-of-Africa migrations.[14] These results were confirmed independently by comparing the fit of competing demographic models with the current population structure of modern *H. pylori* inferred from 60 whole-genome sequences[51]; the best-fitting model involved a diversification of African lineages which predated coalescent events of other lineages, further supporting the hypothesis that *H. pylori* coevolution with human occurred long before their worldwide dissemination.

Using Bayesian clustering analyses performed on concatenated sequences from seven housekeeping gene fragments and one virulence gene, *H. pylori* from a global sample split into seven populations and subpopulations characterized by clear geographical distributions reflected in their name: hpAfrica1 subdivided into two subpopulations, hspWAfrica and hspSAfrica, hpAfrica2, hpEastAsia containing hspAmerind, hspEAsia, and hspMaori subpopulations, and hpEurope.[39] Later on, three additional populations, hpAsia2 and hpNEAfrica[50] and hpSahul[14] were described using extended datasets. All these populations and subpopulations derived from six ancestral populations (Ancestral Africa 1, Ancestral Africa 2, Ancestral East Asia, Ancestral Europe 1 [AE1], Ancestral Europe 2 [AE2], and Ancestral Sahul[14,39,50]). The geographical distribution and genetic relationships between these populations are consistent with the classical model of human migrations, that is, two subsequent waves of migration from Africa into Asia and Europe and a colonization of America from Asia through the Bering Strait and, more recently, from Europe[39] (Fig. 7.1).

The division of the hpEurope population into subpopulations failed. This is probably due to its complex history, namely colonization by several independent waves of migration of genetically distinct populations. This is supported by the observation of two opposite clines, namely AE1 and AE2 within European populations that correlated with the first two principal components of European human diversity.[39,50] Apart from the hybrid genomic structure of hpEurope strains which might have obscured phylogeographical analyses, difficulties in separating the hpEurope population into well-defined clades could also be linked to a lack of phylogenetic signal due to its recent emergence compared to the other *H. pylori* lineages. Interestingly, the genomic analysis of ancient *H. pylori* DNA, extracted from the stomach of the ice-conserved mummy of a European inhabitant estimated to have lived 5300 years before present, did not reveal the typical AE1/AE2 hybrid structure of modern hpEurope strains.[52] Based on 7-loci MLST and whole-genome comparisons, this ancient *H. pylori* genome

Downloaded From Vet-eBooks.com

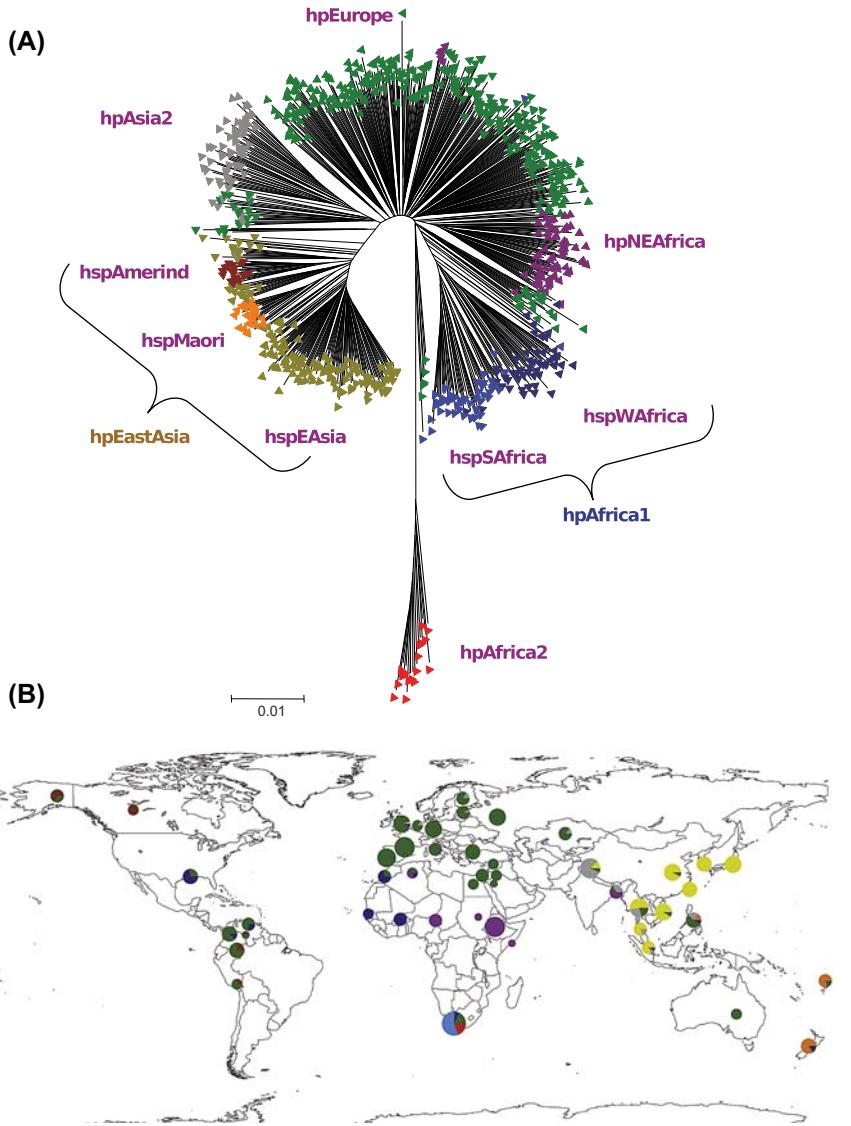


Figure 7.1 (A) Neighbor-joining tree of 769 concatenated sequences from *H. pylori* strains. Colors indicate the population assignment of each strain into one of the nine (sub)populations defined by the Bayesian clustering analysis in Ref. Linz et al.[50]: hpAfrica, hspSAfrica, hspWAfrica, hpNEAfrica, hpEurope, hpAsia2, hspAmerin, hspMaori and hspEasia. (B) Geographical distribution of the nine *H. pylori* (sub)populations. At each sampling location, the proportion of strains assigned to different bacterial (sub)populations are represented by pie charts.
From Linz B, et al. An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* 2007;**445**:915−18.

Downloaded From Vet-eBooks.com

was genetically closest to extant hpAsia2 strains (descendants of AE1 strains) and contained less than 15% of the AE2 component of modern hpEurope strains. This single observation did not allow to draw conclusions about the emergence of hybrid hpEurope strains; however, if the studied strain was representative of *H. pylori* in Copper Age Europe, then its genomic content suggested that the progressive introgression of AE2 strains into Europe was posterior to the Copper Age, being thus much more recent than the divergence times of other *H. pylori* lineages.

Collectively, these studies have shown that geographical and genetic structures of *H. pylori* populations at a global scale were consistent with ancient human migrations. These results were congruent with the accepted human history scenario but they did not necessarily supply more information than former human genetic studies. However, the seminal study of Falush et al.[39] on *H. pylori* populations clearly showed that its spread could be attributed to human migrations. They initiated a series of studies that provided evidence of the usefulness of this bacterium species to infer global but also small-scale evolutionary history of its host. We describe hereafter the main results gathered from some of these studies that improved our understanding of human population structure. Most of these studies are based on the same seven housekeeping gene fragments and implemented their own data in addition of those available on the *H. pylori* MLST database. Unless stated otherwise, the results we relate in the following are based on these gene fragments, Bayesian clustering methods and phylogenetic analyses.

One of the most devious human settling is that of the Pacific. Several scenarios have been suggested depending on the evidence (archaeological, linguistic, or genetic; see Gray et al.[53] for a description of the main models) but the details have remained unclear. As reported in 2009, *H. pylori* isolates from native inhabitants in Taiwan, Papua New Guinea, New Caledonia, and Australia allowed the clarification of Pacific settlement and supplied proof of the utility of this bacterium species to infer host history.[14] This study advocated for a Pacific peopling scenario consisting of two major waves of migration: the first one, from Asia to New Guinea and Australia, which was accompanied by hpSahul strains and occurred 31,000−37,000 years ago, and the second one from Taiwan through the Pacific 5000 years ago, which led to the Austronesian expansion and hspMaori dispersal (Fig. 7.2). Interestingly, these results are consistent with another study that aimed at testing Austronesian expansion using language phylogenies.[53,54] With a large dataset based on language similarities (400 Austronesian languages) and Bayesian phylogenetic methods, Gray et al.[53] resolved the peopling of the Pacific by Austronesian speakers. Like Moodley et al.[14] in the genetic study, Gray et al.[53] observed that Austronesian people originated from Taiwan about 5200 years ago. The linguistic study also described several expansion pulses and pauses after the first migration from Taiwan that led to the actual distribution of Pacific people (see Ref. 53; Fig. 7.2).

Another attractive aspect of *H. pylori* is that its population structure reflects human history at a local or fine scale.[15,50,55] For instance, Wirth et al.[15] were able to detect differences in population genetic structure of *H. pylori* from Ladakhi Buddhists and Muslims, two major ethnic groups in Ladakh socially separated in this province since 500−1000 years due to cultural and religious differences. The *H. pylori* isolates from

Downloaded From Vet-eBooks.com

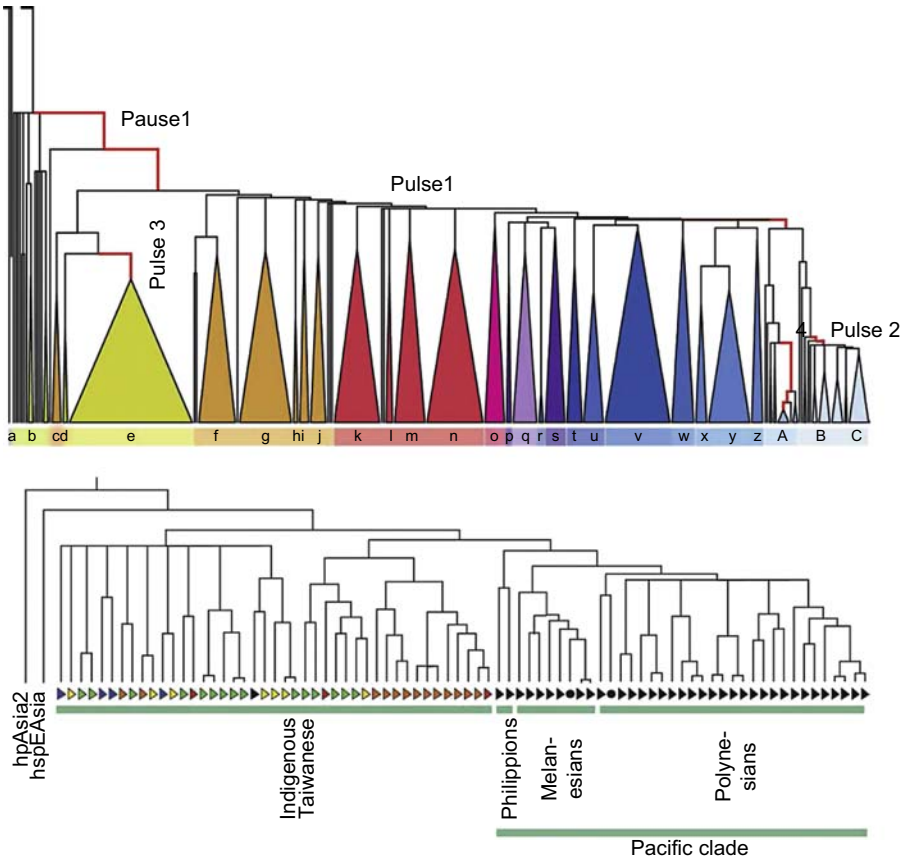


Figure 7.2 The peopling of the Pacific Islands from human bacterial (at the bottom) and language phylogenies (at the top).
From Renfrew C. Where bacteria and languages concur. *Science* 2009;**323**:467−68.

the two ethnic groups presented different ancestries: while isolates from Buddhists derived from AE1 and Ancestral East Asia populations, isolates from Muslims mostly derived from AE1 and a few strains from AE2. Therefore, Buddhism was introduced in Ladakh by Tibetans (carrying hpEastAsia bacteria) into an ancestral Ladakhi population (carrying AE1 bacteria) and Islam by a few people carrying AE2 bacteria.[15] Altogether, these results reflect cultural differences and the recent history of population movements in Ladakh which were not detectable with classical human markers (microsatellites and the mitochondrial D-loop). In the same vein, Latifi-Navid et al.[55] unraveled genomic differences between *H. pylori* isolates from Iranian patients, which correlated with ethnic groups and geographical locations. Using the Bayesian clustering method, Iranian *H. pylori* isolates fell into the hpEurope population and derived from AE1 and AE2. However, at the population level, most of Iranian and European populations were genetically differentiated. A neighbor-joining tree based on

Downloaded From Vet-eBooks.com

pairwise F_{ST} revealed that *H. pylori* isolated from the different Iranian populations were not clustered together but clustered with strains isolated from nearby (ethnically or geographically) countries reflecting historical contacts during the past 2000 years.[55] Thus, the Iranian-Arab population, which reached Iran during the Islamic conquest of Persia in the 7th century, was close to the Palestinian and Israeli isolates; two populations from West Central Iran were close to Turkish strains, probably reflecting contacts with Turks during the Ottoman Empire and later; and two populations from the north eastern part of Iran were close to isolates from Uzbekistan, which could be the consequence of contacts during the fight for the control of the region between Uzbeks and Iranians in the 16th century.[55]

Other studies focused on even finer scales. Either within areas that have complicated history because of their localization, at the boundaries of several continents, or where multiple ethnic groups were present. For instance, investigations on *H. pylori* isolated from the three major ethnic groups in Malaysia (Chinese, Indians, and Malays), revealed a common origin of Malaysian Indian and Malay strains, which was different from that of Chinese isolates.[56] The authors observed that most Indian isolates clustered within the hpAsia2 population and most Chinese within the hspEAsia population. In contrast, Malay strains were not assigned to one particular population but belonged to five different (sub)populations, the majority of them belonging to the hspAsia2 one. By completing the hpAsia2 population with Malaysian isolates, the authors identified a subdivision of this population into two subpopulations, hspLadakh which contained Ladakhi strains only, and hspIndia comprising the majority of both Malaysian Indian and Malay isolates. Tay et al.[56] advocated for a common origin of Malaysian Indian and Malay strains which was different from that of Chinese isolates and a recent acquisition of *H. pylori* by Malay populations from other populations.

Devi et al.[33] detected an hpEurope genetic signature for the majority of strains from native Peruvians but also detected an hpAmerind signature for some of them ($\sim$20%). At the same time, the analysis of the *cag*PAI revealed that all native Peruvian strains, even the hpAmerind ones, presented a "western" type. The authors concluded that European strains that have spread into South America during the past 500 years might have outcompeted the hpAmerind ones probably as a result of a selective advantage. In addition, they suggested than lateral gene transfer occurred between hpEuropean and hpAmerind strains during the colonization of Peru since some isolates presented two different genetic signatures depending on the markers (housekeeping genes or *cag*PAI).

Finally, Devi et al.[57] observed that 63 Indian isolates showed European genetic signatures, suggesting a common ancestral origin between the two populations. The authors suggested that *H. pylori* strains might have arrived with the Indo-Europeans about 4,000−10,000 years ago.

To conclude, *H. pylori*, with its unexpected high diversity has been proven to be a good human migration tracer, both on short and long time scales. Despite frequent recombination events, its populations were geographically structured suggesting that recent migrations did not completely obscure the signatures left by geographical isolations and therefore rather reflect ancient human history.[58] Moreover, the presence of recombination allows the detection of admixture between several ancestral populations,

Downloaded From Vet-eBooks.com

revealing multiple independent waves of migration and sometimes multiple ethnicity signatures within a single genome. Today, the frequency of infections has decreased, in industrialized countries in particular (in United States, less than 10% of the children are infected[59]), highlighting the need for other candidates to infer human evolutionary history and to urgently collect specimens from endangered ethnic groups.

3.1.2 *Mycobacterium tuberculosis*

Mycobacterium tuberculosis has plagued mankind for millennia[62] and continues to do so, with a worldwide death toll for tuberculosis that reached 1.5 million for the year 2014 (WHO, 2015). This Gram-positive bacterium belongs to the *M. tuberculosis* complex (MTBC) which includes seven other closely related species and subspecies infecting both humans and animals. Each (sub)species of the MTBC shows a distinct host preference without being dependent of this sole host.[63] Airborne transmission of tubercle bacilli involves their excretion in droplets that can be inhaled and penetrate the lung. *M. tuberculosis* infects one-third of the world population although prevalence and mortality are higher in developing countries. Most infections are latent, but 5−10% of the infections expand into disease.[63]

The MTBC presents a strictly clonal population structure with none or few recombinations.[64−67] Initial studies of MTBC population reported low rates of genetic diversity and weakly resolved or star-like phylogenies; however, these studies had been hampered by important technical limitations either due to the choice of the markers and/or to problems linked with ascertainment bias.[68,69] Advances in mycobacterial genomics have, however, revealed higher levels of genetic diversity than was previously thought and documented the presence of geographical and/or ethnical structures within *M. tuberculosis* populations.[68,70] These include studies based either on genes sequences[69,71,72] or on mycobacterial interspersed repetitive units (MIRUs that contained variable number of tandem repeats (VNTRs)[67,73,74]).

MIRU-VNTR markers combined with an extensive dataset allowed to tackle the origin, timing, and spread of the MTBC.[67] The authors drew a *M. tuberculosis* phylogeny from 24 MIRU loci and from 355 isolates representative of the MTBC distribution and estimated the divergence time between main clades using two approaches, a probabilistic and a distance-based one. Both individual and population-based phylogenies evidenced two major lineages that were confirmed by a Bayesian clustering analysis. These lineages distinctly separated *M. tuberculosis sensu stricto* strains (clade 1), all from humans with the exception of East African Indian (EAI) population, from the animal-infecting strains, and the West African strains (clade 2). EAI strains were basal in the phylogeny, suggesting a human origin for the animal-associated clade. Clade 1 presented a geographical substructure with African, Asian, Latin American-Mediterranean, and African-European clusters. In the same study, the mutation rate of VNTR loci was estimated to about 10^{-4} change per locus per year. Interestingly, the emergence of the two major clades was dated at about 40,000 years ago, consistent with the initial human migrations out of Africa. Clade 1 emerged from the MTBC about 30,000 years ago and dispersed with humans through the other continents through several waves of migration. Clade 2 dated at about 20,000 years and

Downloaded From Vet-eBooks.com

descended from an EAI-like population that has most likely been transmitted from humans to animals (and not the other way around) in the Fertile Crescent about 13,000 years ago when domestication began. Based on Bayesian analyses, all human *M. tuberculosis* populations exhibited consistent expansion rates, with the highest expansion, a 500-fold increase, detected for the Beijing population. Beijing lineage expansion began 180 years ago, concomitant to the modern demographic explosion of humans, the industrial era, and modern intercontinental movements. Hence, this study highlights the noteworthy parallel demographic evolution between humans and *M. tuberculosis*.

Whichever genetic markers are used (large sequence polymorphisms, SNPs, indel analyses), global phylogenies were concordant and led to a biogeographical consensus with six lineages associated to particular areas which may reflect human demographic history.[68,70] For instance, Hershberg et al.[69] constructed maximum parsimony and neighbor-joining phylogenies of MTBC using 89 concatenated gene sequences from 108 strains comprising a representative sample of the MTBC. Both phylogenies were congruent and their topologies were comparable to the phylogeny from Wirth et al.[67] including the two primary branches splitting "ancient" and "modern" lineages according to Brosch et al.[75] (Fig. 7.3A). The presence of the six main lineages in Africa and the deeply rooted West African branches (the only deeply rooted ones) argued in favor of an African origin of *M. tuberculosis* and its sequential spread into Europe and Asia. These hypotheses were fully confirmed and refined by larger-scale studies based on whole-genome sequences. In such a study, Comas et al.[62] used a panel of 259 MTBC genomes to reconstruct the evolutionary history of the complex (Fig. 7.3A and B). MTBC phylogeny was well resolved and showed large-scale similarities with the topology of human mitochondrial DNA phylogeny; major clades in both the MTBC and human phylogenies were superimposable and shared geographical distributions, most notably between MTBC lineage 1 and human macrohaplogroup M, both distributed in Southeast Asia and Oceania, or between MTBC lineages 2−4 and human macrohaplogroup N, both distributed in Eurasia (Fig. 7.3C and D). These striking similarities indicated coevolution and codivergence of MTBC lineages with their human hosts over long time scales. Comparisons of demographic models indicated that MTBC infected humans since at least about 70,000 years ago and before the out-of-Africa migrations.

Therefore, *M. tuberculosis* evolution reflects past human history over long time scales. Additionally, it appears to mirror recent colonization movements and demographic changes in human populations. For instance, Euro-American strains were almost ubiquitous, possibly reflecting the numerous European migrations from Europe to America during the 19th century, and to Africa, Asia, and Middle East during post-Columbian era. Similarly, the presence of East Asian strains in South Africa might be explained by the import of slaves from Southeast Asia by Dutch colonialists during the 17th and 18th centuries or by Chinese migrants who came into South Africa in early 1900s to work in gold mines.

MTB therefore appears to be of particular interest in the inference of recent host history, which has been best illustrated in studies focusing on the highly successful MTBC Beijing lineage, which has gathered much attention due to its propensity to

Downloaded From Vet-eBooks.com

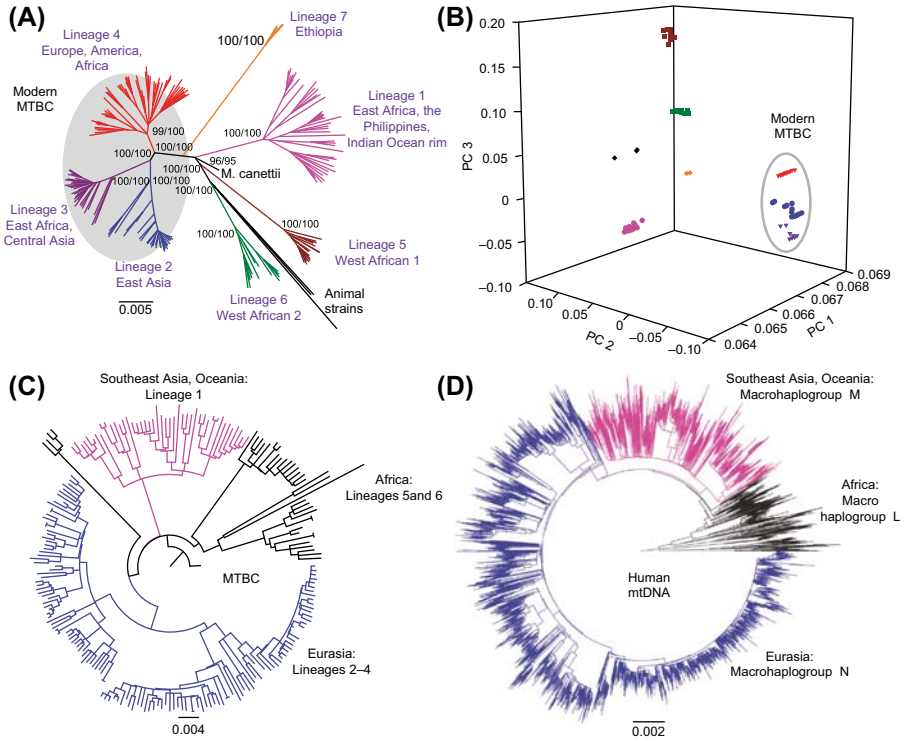


Figure 7.3 The genome-based phylogeny of MTBC mirrors that of human mitochondrial genomes. (A) Whole-genome phylogeny of 220 strains of MTBC. Support values for the main branches after inference with neighbor-joining (left) and maximum-likelihood (right) analyses are shown. (B) Principal-component analysis of the 34,167 SNPs. The first three principal−component axes (PC 1−PC 3) are shown; these discriminate between evolutionarily modern (*gray circle*) and ancient (all other) strains. Individual lineages are shown with the same colors as in (A). (C and D) Comparison of the MTBC phylogeny (C) and a phylogeny derived from 4955 mitochondrial genomes (mtDNA) representative of the main human haplogroups (D). Color coding highlights the similarities in tree topology and geographical distribution between MTBC strains and the main human mitochondrial macrohaplogroups (black, African clades: MTBC lineages 5 and 6, human mitochondrial macrohaplogroups L0−L3; pink (gray in print versions), Southeast Asian and Oceanian clades: MTBC lineage 1, human mitochondrial macrohaplogroup M; blue (dark gray in print versions), Eurasian clades: MTBC lineage 2−4, human mitochondrial macrohaplogroup N). MTBC lineage 7 has only been found in Ethiopia, and its correlation with any of the three main human haplogroups remains unclear. Scale bars indicate substitutions per site.
From Comas I, et al. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat Genet* 2013;**45**:1176−82.

harbor drug-resistant strains. Mokrousov[74] collected the data of 11 VNTR loci in 1302 Beijing strains, mainly from Eurasia, and performed phylogenetic network and multidimensional scaling (MDS) analyses. He observed that the geographical distribution of this particular *M. tuberculosis* lineage in Eurasia mirrors geographical, political, and

Downloaded From Vet-eBooks.com

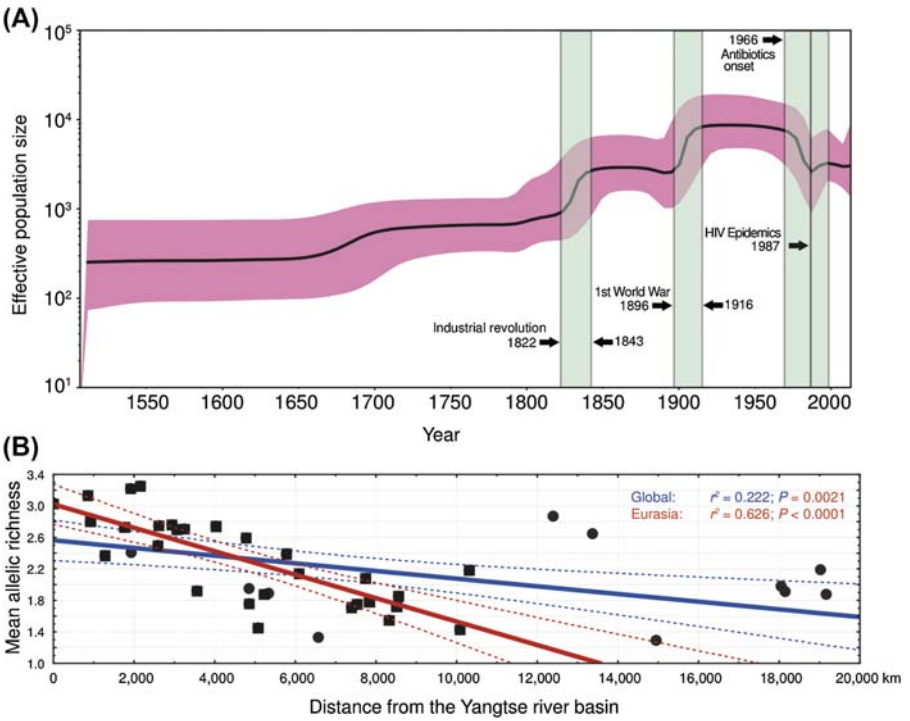


Figure 7.4 (A) Bayesian skyline plot indicating changes in the Beijing lineage over time with a relaxed molecular clock set at 1×10^{-7} mutations per nucleotide per year. The *shaded area* represents the 95% confidence intervals, and the *green colored boxes* (light gray in print versions) represent major socioeconomic events that might have affected the demography of *M. tuberculosis*. (B) Genetic erosion out of China. Mean allelic richness within geographical populations is plotted against geographical distance from the Yangtze River basin. *Filled squares* denote the Eurasian samples used for the regression; *filled circles* correspond to the global collection. Confidence intervals are represented by *dashed lines*.
From Merker M, et al. Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nat Genet* 2015;**47**:242−49.

sociocultural differences.[74,76] More detailed insights into the population structure of the Beijing lineage at larger scales were obtained in a later study combining 24-loci MIRU-VNTR in almost 5000 Beijing strains and whole-genome sequences in a subsample of 110 representatives.[77] Using Bayesian approaches, several historical events could be linked to abrupt changes in the effective population size of the Beijing lineage. A concomitant increase of tuberculosis-effective population size was observed with the industrial revolution and World War I, followed by a sharp decrease concomitant to the introduction of effective antibacterial chemotherapies (Fig. 7.4A). The Beijing lineage as a whole could be subdivided into clonal complexes (CCs) whose geographical distribution and demography history closely reflected immigration episodes in Asia, the cradle of the lineage. For instance, population expansions of CCs

Downloaded From Vet-eBooks.com

1 and 2 during the late 19th century in southern regions of the Russian empire could be linked to waves of Chinese migration toward Kazakhstan and Uzbekistan from 1861 to 1877. Furthermore, the routes out of China left a typical genetic erosion signature, with the highest genetic diversity observed in the Yantze River region and the lowest on other continents (Fig. 7.4B).

Another study, that focused on Taiwan, documented associations between *M. tuberculosis* genotypes and ethnic and migratory populations.[73] The authors studied strains from three Taiwanese populations differing in their ethnicity and in the time of arrival in Taiwan: the first population was composed of aborigines who descended from Austronesian people and inhabited Taiwan at least since the 16th century, while the others were two Han Chinese populations. The two latter populations were composed of veterans who were born in Mainland China and migrated into Taiwan during the Chinese civil war during 1946−1950, and the general Taiwanese population composed of individuals whose ancestors migrated from China 200 to 400 years ago. Hence, the three populations differed in their ethnicity and/or in their time of arrival in Taiwan. The authors discriminated the different genotypes by combining several genetic tools (VNTR, spoligotyping, and indel analyses) which provided a high discriminatory power. They performed a UPGMA tree and looked at the separation of the genotypes. Their analyses showed that the genotypes differed between populations and that isolates from aborigines were comparatively more ancient than those from veterans and from the general Taiwanese population, thus arguing in favor of an association between genotypes, ethnicity, and migratory movements (i.e., the length of migratory time in Taiwan[73]).

Hence, as in the case of *H. pylori, M. tuberculosis* seems to harbor a population-genetic structure, even at a small scale, that correlates with human history. Past and recent human migrations and expansions influenced *M. tuberculosis* population structure probably because of the strong and stable association between the two protagonists.[64,78] Some discrepancies have however been highlighted. This is the case in the study of Mokrousov[74] in which proximities between Japanese and Russian populations were found. Moreover, the pattern observed in Chinese strains was unclear suggesting that *M. tuberculosis* structure might reflect unknown human migrations, epidemiological links between these populations, or that *M. tuberculosis* population structure is affected by high rates of horizontal transmission events during epidemics in these regions, blurring the coevolution signal.

Likewise, for *M. tuberculosis*, it has been shown that *Mycobacterium leprae* reflects human migrations.[79,80] This pathogen causes chronic dermatological and neurological diseases and has humans as a unique known reservoir. The *M. leprae* genome contains an amazing number of pseudogenes and is exceptionally well conserved.[79−81] Like its close relatives, *M. leprae* is highly clonal.[79] Monot et al.[79] first described four subtypes that were geographically structured from three informative SNPs (SNP type1 were found in Asia, pacific, and East Africa, SNP type 4 in West Africa and the Caribbean, SNP type 3 in Europe, North Africa, and Americas, and SNP type 2, the rarest, in Ethiopia, Malawi, North India, and New Caledonia). They proposed a scenario for the origin of leprosy and its spread though ancient human migrations, colonization, and the slave trade. However, the origin of leprosy was unclear. In a second study

Downloaded From Vet-eBooks.com

with an extended dataset, they confirmed their initial results and clarified the origin of leprosy. According to these two studies, SNP type 2 from South Africa was the ancestral type and gave rise to SNP type 1 and SNP type 3 which dispersed into Asia and Europe, respectively. However, two independent introductions of leprosy seemed to occur in Asia: the first one occurred when humans left Africa and a second one came from Europe and was associated to the Silk Route. SNP type 4 appeared to originate from SNP type 3 and was found in West Africa and countries linked to the slave trade. Finally, the authors observed that leprosy in America most likely originated from Europe rather than from Asia though the Bering Strait reflecting the relatively recent massive migrations from Europe to America. In this context, *M. leprae* genotypes have been shown to correlate with patient's ancestry in the Colombian population, whose structure reflects migration of Europeans and Africans in a Native American population.[82] The two major *M. leprae* haplotypes present in the population, namely C54 and T45, are associated with African and European origins of the patients, respectively, which strongly suggests coevolution of *M. leprae* and its human host long before the colonization of Colombia.

3.2 Viruses

Many viruses have been proposed as providing valuable insights into human population history. Among them we can cite human T-cell lymphotrophic viruses 1 and 2 (HTLV-1 and -2), the human polyomavirus JC (JCV) and its closely related BK virus (BKV), the human papillomavirus (HPV), the herpes simplex virus (HSV),[83] and the hepatitis G virus (GBV-C/HGV). However, most often, the hypothesis of viral and human codivergences is not well supported or evidenced and/or is founded only on geographical associations which may be coincidental or may result from other factors such as natural selection (see Ref. Wirth et al., 16). Indeed, most of these viruses seemed to suffer from drawbacks making them poor candidates to elucidate past human migrations.[19,84] For instance, a majority of viruses are often transmitted horizontally which leads to fast genetic admixture. Here, we will detail the case of JCV to illustrate some of the kinds of problems we can encounter when studying viruses with regard to human history.

3.2.1 The Human Polyomavirus

One of the most investigated viruses in the context of human migrations is the human polyomavirus JC. JCV is a double-stranded DNA virus which is responsible for harmless kidney infections, except for immunocompromised patients where leukoencephalopathy can develop.[85] The virus is acquired during childhood and persists in renal tissues for life.[86] JCV is human-specific and ubiquitous with an estimated seroprevalence of at least 70% in the human population.[87] Evidence for both vertical and horizontal transmission have been found although horizontal transmission seems to occur preferentially between closely related populations.[88–90] In addition, Kitamura et al.[91] detected identical strains from sequential samples from the same patients taken about 6 years apart suggesting that multiple infections might rarely occur, thus limiting

Downloaded From Vet-eBooks.com

recombination events between divergent JCV strains. All these features have led to the hypothesis that JCV might be useful to reconstruct human migrations.

Several major viral strains and subtypes showing geographical associations have been described from partial gene[92,93] and whole-genome sequences.[94] Since then, numerous studies at global and local scales have documented the genetic relationships and the geographical distributions of these genotypes. Briefly, type 1 was found mainly in Europe,[95,96] types 2 and 7 in Asia,[92,93,96,97] types 3 and 6 in Africa,[96,98,99] type 4 in the USAs and Europe,[95,100] and type 8 has been detected in Papua New Guinea and the Pacific Islands.[101,102] Type 2 was subdivided into several subtypes presenting variations in abundance according to area: subtype 2A was preponderant in the Japanese and Native American populations, subtype 2B in Eurasians, subtype 2D in Indians, and 2E in Australians and populations from the West Pacific.[102] In the same way, type 7 included subtype 7A preponderant in Southern China and South-East Asia, and subtype 7B which was found in higher proportion in Northern China, Mongolia, and Japan.[103] Cui et al.[103] detected a third subtype called 7C in northern and southern China. Finally, type 5 was shown to combine type 6 and type 2B sequences and is the unique example of recombinant JCV strain.[104]

Interestingly, the multiple origins of American people were detectable by analyzing JCV genotypes.[100] Native Americans represented by two ethnic groups (Flathead People and Navaho) were mostly infected by subtype 2A, a genotype found in East Asia and Japan, which may reflect an Asian origin through the Bering Strait.[92] In contrast, European Americans carried type 1 (European genotype) for a majority and in lower proportion type 4 (14%) and types 2 (less than 10%).[100] Surprisingly, no type 6 was found in the African-American population but type 1 (32%), type 4 (44%), and type 3 (18%) were found in them, suggesting a genetic admixture between African and European types and reflecting both past and recent migratory movements.[105] Stoner et al.[100] suggested that the high frequency of European strains in African European populations could be due to a selective advantage of these strains compared to African ones.

JCV populations also appear to be geographically structured in the Pacific Islands, probably due to multiple human migration waves.[101,102] Four subtypes were identified within JCV populations from western Pacific Islands: subtype 8A restricted to Papua New Guinea, subtypes 8B from non-Austronesians, 2E from Austronesians widely distributed though Pacific Islands, and subtype 7A rarely found. Yanagihara et al.[102] proposed that subtype 8A first arrived in Papua New Guinea or Sahul followed by subtype 8B. Later ($\sim$5000 years ago), Austronesian expansion might have led to the spread of subtype 2E. Recent migrations from South China or Taiwan might have brought subtype 7A into Guam. Surprisingly, Australian JCV strains belonged to subtype 2E which is genetically close to the subtype found in East Asia (subtype 2A). This is in sharp contradiction with the known history of Pacific peopling which was confirmed by *H. pylori* population studies and language phylogeny (see Section 3.1.1). Indeed, the first wave of migrations from Asia into the Pacific Islands led to the peopling of both Australia and New Guinea. Therefore, we expected to find in native Australian strains the same subtypes as those found in New Guinea.

Downloaded From Vet-eBooks.com

In accordance with these results, Pavesi[106–108] tackled the evolution of JCV genotypes by means of principal component analyses based on JCV sequences from the five continents. These analyses evidenced that the African type 6 might be the ancestral genotype that gave rise to two major independent lineages, one clustering types 1 and 4 while the other containing types 2, 3, 7, and 8. This analysis has led the author to propose an alternative to the classical model of human migrations, namely "the two-migration model." This model hypothesizes two early routes of expansion out of Africa: one route into Asia and the second one into Europe. However, phylogenies based on whole JCV−genome sequences showed some discrepancies.[109] For instance, the basal European clade position was paradoxal.[93,95,101,104] This is inconsistent with the hypothesis of an infection of humans by JCV before their expansion from Africa. Pavasi[106] handled this question by reconstructing two phylogenies based on slow- and fast-evolving sites defined from the Shannon entropy index. Phylogenies based on invariants plus slow-evolving sites and on invariants plus fast-evolving sites were different. When invariants and slow-evolving sites were used to reconstruct phylogeny, the topology was similar to topologies obtained from the whole-genome sequences with the European clade at the basal position and type 6 as the ancestral type of all other types. In contrast, the phylogeny based on invariants and slow-evolving sites placed the type 6 on the deepest branch. This is consistent with an African ancestry. However, other questionable findings remain to be clarified such as the higher genetic diversity observed in European and Asian than in African JCV.[109] Coincidences between geographical association between JCV and human populations may result from other factors such as natural selection or specific viral life-history traits. More studies on JVC are therefore needed before concluding with regards to human migrations. In addition, the molecular clock needs to be carefully reevaluated.[109]

One debatable point of all these studies is that they have relied on the hypothesis of JVC and human codivergence and on a slow mutation rate which has not been tested independently from the coevolution hypothesis. Mutation rates were first estimated to range between 10^{-7} and 4×10^{-7} per site per year.[104,110] These estimations were founded on the assumption of a longtime coevolution between JCV and humans (at least since the expansion from Africa about 150,000 years ago) and estimations were calibrated from host divergence times. Hence, this approach is somewhat tautological. In contradiction with the preceding, two more studies found much faster mutation rates using a Bayesian Markov chain Monte Carlo (MCMC) approach which is free from the assumption of codivergence and is based on coalescent analysis of sequences sampled at different times.[111,112] Shackelton et al.[112] tested congruence between JCV and human phylogenetic trees by mapping consensus JCV trees onto three possible human trees thus creating "tanglegrams." From each of these tanglegrams, the potentially optimal solutions were determined by evaluating the noncoevolutionary events (i.e., duplication, horizontal transfer, and loss of a virus by a host population) required to reconcile JCV and human trees. Randomizations of the branches of the viral tree were used to test the hypothesis that the viral tree was more congruent with the human tree than a random tree would be. In both studies, no evidence for codivergence between human and virus phylogenies was found.[111,112] Shackelton et al.[112] estimated for humans the age of the MRCA to be in accordance with the accepted estimates (i.e., between

Downloaded From Vet-eBooks.com

100,000 and 150,000 years) and provided evidence for an expansion starting 50,000 year ago when major cultural changes occurred. In contrast, the MRCA for JCV was estimated not to exceed 3100 years ago. Both studies found a significantly higher mean substitution rate for JCV than previous estimations (more than 100-fold faster: 1.7×10^{-5} see Ref. 112 and 3.6×10^{-5} substitutions/site/year see Ref. 111). Considering this faster mutation rate, skyline plots, a coalescent method for estimating past population dynamics,[113] revealed that the global viral population increased during the last 350 years[112] and that posterior population estimates for viruses and humans differed totally (Fig. 7.5; see Refs. 111,112).

These last two studies demonstrated that JCV populations should not be used to infer past human history because their population dynamics occurred at time scales that are

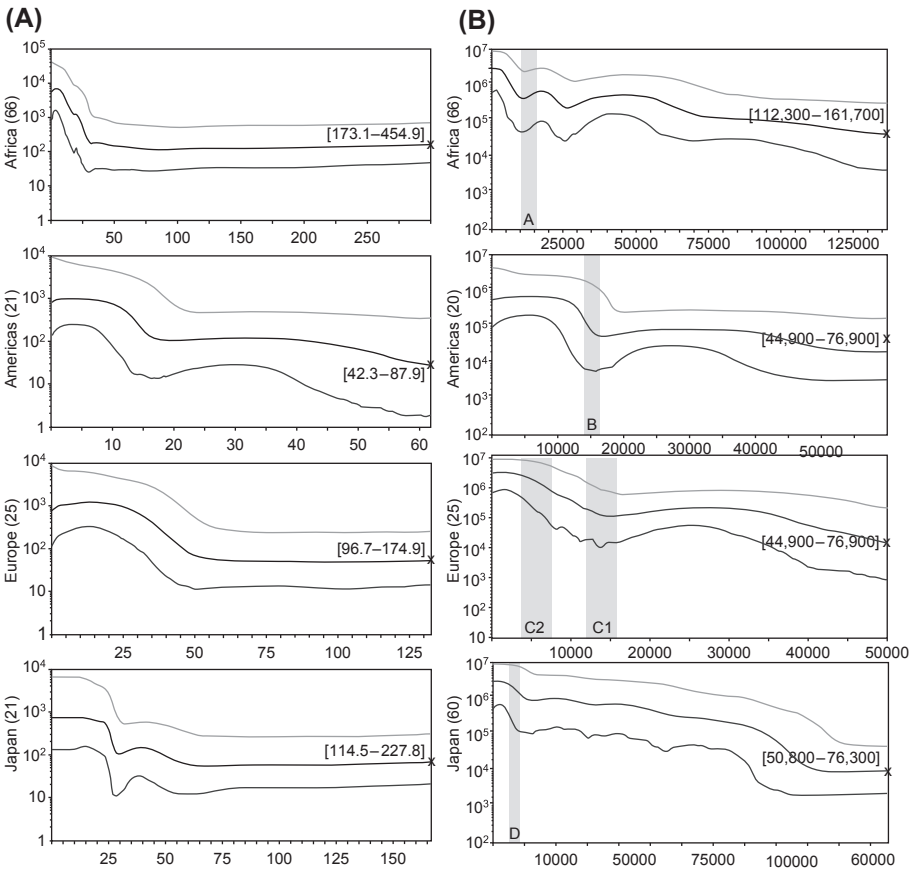


Figure 7.5 Bayesian skyline plots performed on JCV genomes (A) and on human mtDNA (B) sequences with nucleotide substitution rates of 3.64×10^{-5} and 1.7×10^{-8} for virus and human sequences, respectively. The *x*-axis is the number of years before present and the *y*-axis is the scaled population size ($=N_e \times g$). The median estimate of the population size ($N_e \times g$; *black line*) and the 95% confident intervals (*light gray line*) are given.
From Kitchen A, Miyamoto MM, Mulligan CJ. Utility of DNA viruses for studying human host history: case study of JC virus. *Mol Phylogenet Evol* 2008;**46**:673–82.

Downloaded From Vet-eBooks.com

too recent. It seems more likely that JCV population phylogenies and dynamics reflect recent societal and epidemiological shifts in human behavior or technological innovations.[111] In agreement with this, skyline plots indicated expansions of JCV in Africa, Europe, and Japan that began about 50 years ago.[111] Expansions in Europe and Japan 50 years ago may be due to societal changes that occurred after the World War II.

4. Conclusion

In this chapter, we sketched a brief overview of the emerging field of the use of microbes or parasites as proxies for human migrations. Now, in 2016, more than two decades after the first attempts using viruses for this purpose, we have reached the age of maturity. Though many candidates have been tested, the most convincing and innovative results were obtained with *H. pylori*. This was mainly driven by the intrinsic qualities of this bug, but also by a decent sampling scheme, the accumulation of large data sets, and a good geographical coverage. Indeed, *H. pylori* MLST brought a lot of information in terms of genetic diversity and structure of this pathogen and indirectly of its human host. However, with the continuous improvement of sequencing technologies and the increasing facility to generate complete microbial genome at reasonable costs, the paradigm of *H. pylori* MLST might shift toward other bacterial species that displayed too few mutations in an MLST scheme but might unravel valuable information at the genomic scale. The ongoing shift from a multiple gene approach to a genomic approach is paving the way toward a new golden age in this field, which will increase the demand for new population genomic tools and algorithms. The next targets will certainly belong to the group of relevant human pathogens whose study is more easily supported, but commensals should not be dismissed. For example, to circumvent the complicated sampling of *H. pylori* bacteria, Henne et al.[114] proposed the use of bacteria from saliva samples to trace human migration. The data obtained from human microbiota metagenomics could also be of use in this context.[115] The evolutionary history of our microscopic "companions" can be seen like a multilayer information box; different species might not provide us with a unique congruent scenario, but might instead unravel the complexity of host–parasite interactions from neutral coinciding genetic patterns to extreme selection biases. Extending the use of bugs to infer host migrations of other mammals or organisms is one more future challenge we might face. The origin and dispersal of our animal stocks could be revisited; the demography and "ghost" genetic structure of endangered species (big cats) could be evaluated. Overall, the limits today are our imagination and the difficulty to handle these new-generation, large-scale metadata.

Abbreviations

AE1	Ancestral Europe 1
AE2	Ancestral Europe 2
BKV	BK polyomavirus
***cag*PAI**	*cag* pathogenecity island

Downloaded From Vet-eBooks.com

EAI	East African Indian
GBV-C/HGV	hepatitis G virus
HPV	human papillomavirus
HTLV	human T-cell lymphotrophic virus
HVR	hypervariable region
IBD	isolation by distance
IGSR	intergenic spacer region
JCV	JC polyomavirus
LCR	long control region
MCMC	Markov chain Monte Carlo
MIRUs	mycobacterial interspersed repetitive units
MLST	multilocus sequence typing
MRCA	most recent common ancestor
MTBC	*Mycobacterium tuberculosis* complex
mtDNA	mitochondrial DNA
SNP	single nucleotide polymorphism
UPGMA	unweighted pair group method with arithmetic mean
VNTR	variable number of tandem repeats

References

1. Cavalli-Sforza LL. The DNA revolution in population genetics. *Trends Genet* 1998;**14**: 60−5.
2. Cavalli-Sforza LL, Feldman MW. The application of molecular genetic approaches to the study of human evolution. *Nat Genet* 2003;**33**:266−75.
3. Kundu S, Ghosh SK. Trend of different molecular markers in the last decades for studying human migrations. *Gene* 2015;**556**:81−90.
4. Fu Q, et al. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr Biol* 2013;**23**:553−9.
5. Poznik GD, et al. Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science* 2013;**341**:562−5.
6. Pagani L, et al. Tracing the route of modern humans out of Africa by using 225 human genome sequences from Ethiopians and Egyptians. *Am J Hum Genet* 2015;**96**:986−91.
7. Eshleman JA, Malhi RS, Smith DG. Mitochondrial DNA studies of Native Americans: conceptions and misconceptions of the population prehistory of the Americas. *Evol Anthropol* 2003;**12**:7−18.
8. Hurles ME, Matisoo-Smith E, Gray RD, Penny D. Untangling Oceanic settlement: the edge of the knowable. *Trends Ecol Evol* 2003;**18**:531−40.
9. Henn BM, Cavalli-Sforza LL, Feldman MW. The great human expansion. *Proc Natl Acad Sci USA* 2012;**109**:17758−64.
10. Liu H, Prugnolle F, Manica A, Balloux F. A geographically explicit genetic model of worldwide human-settlement history. *Am J Hum Genet* 2006;**79**:230−7.
11. Prugnolle F, Manica A, Balloux F. Geography predicts neutral genetic diversity of human populations. *Curr Biol* 2005;**15**:R159−60.
12. Ramachandran S, et al. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc Natl Acad Sci USA* 2005;**102**:15942−7.

Downloaded From Vet-eBooks.com

13. O'Rourke DH, Raff JA. The human genetic history of the Americas: the final frontier. *Curr Biol* 2010;**20**:R202−7.
14. Moodley Y, et al. The peopling of the Pacific from a bacterial perspective. *Science* 2009; **323**:527−30.
15. Wirth T, et al. Distinguishing human ethnic groups by means of sequences from *Helicobacter pylori*: lessons from Ladakh. *Proc Natl Acad Sci USA* 2004;**101**:4746−51.
16. Wirth T, Meyer A, Achtman M. Deciphering host migrations and origins by means of their microbes. *Mol Ecol* 2005;**14**:3289−306.
17. Whiteman NK, Parker PG. Using parasites to infer host population history: a new rationale for parasite conservation. *Anim Conserv* 2005;**8**:175−81.
18. Nieberding CM, Olivieri I. Parasites: proxies for host genealogy and ecology? *Trends Ecol Evol* 2007;**22**:156−65.
19. Holmes EC. The phylogeography of human viruses. *Mol Ecol* 2004;**13**:745−56.
20. Rannala B, Michalakis Y. Population genetics and cospeciation: from process to pattern. In: Page RDM, editor. *Tangled trees: phylogeny, cospeciation and coevolution*. University of Chicago Press; 2003. p. 120−43.
21. Schwarz S, et al. Horizontal versus familial transmission of *Helicobacter pylori*. *PLoS Pathog* 2008;**4**:e1000180.
22. Schierup MH, Hein J. Consequences of recombination on traditional phylogenetic analysis. *Genetics* 2000;**156**:879−91.
23. Martin DP, Williamson C, Posada D. RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* 2005;**21**:260−2.
24. Martin DP, et al. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 2010;**26**:2462−3.
25. Ward BJ, van Oosterhout C. Hybridcheck: software for the rapid detection, visualization and dating of recombinant regions in genome sequence data. *Mol Ecol Resour* 2016;**16**: 534−9.
26. Gibbs MJ, Armstrong JS, Gibbs AJ. Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 2000;**16**:573−82.
27. McVean G, Awadalla P, Fearnhead P. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* 2002;**160**:1231−41.
28. Nielsen R. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics* 2000;**154**:931−42.
29. Worobey M. A novel approach to detecting and measuring recombination: new insights into evolution in viruses, bacteria, and mitochondria. *Mol Biol Evol* 2001;**18**:1425−34.
30. Maynard Smith J, Smith NH. Detecting recombination from gene trees. *Mol Biol Evol* 1998;**15**:590−9.
31. Jakobsen IB, Easteal S. A program for calculating and displaying compatibility matrices as an aid in determining reticulate evolution in molecular sequences. *Comput Appl Biosci* 1996;**12**:291−5.
32. Achtman M, et al. Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Mol Microbiol* 1999;**32**:459−70.
33. Devi SM, et al. Genomes of *Helicobacter pylori* from native Peruvians suggest admixture of ancestral and modern lineages and reveal a western type *cag*-pathogenicity island. *BMC Genomics* 2006;**7**:191.
34. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009;**25**:1451−2.
35. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 2007;**24**: 1586−91.

Downloaded From Vet-eBooks.com

36. Foll M, Gaggiotti O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 2008;**180**:977−93.
37. Duforet-Frebourg N, Bazin E, Blum MG. Genome scans for detecting footprints of local adaptation using a Bayesian factor model. *Mol Biol Evol* 2014. http://dx.doi.org/10.1093/molbev/msu182.
38. Farris JS, Kallersjo M, Kluge AG, Bult C. Testing significance of incongruence. *Cladistics* 1994;**10**:315−9.
39. Falush D, et al. Traces of human higrations in *Helicobacter pylori* populations. *Science* 2003;**299**:1582−5.
40. Dunn BE, Cohen H, Blaser MJ. Helicobacter pylori. *Clin Microbiol Rev* 1997;**10**:720−41.
41. Kivi M, et al. Concordance of *Helicobacter pylori* strains within families. *J Clin Microbiol* 2003;**41**:5604−8.
42. Suerbaum S, Michetti P. *Helicobacter pylori* infection. *N. Engl J Med* 2002;**347**:1175−86.
43. Suerbaum S, et al. Free recombination within *Helicobacter pylori*. *Proc Natl Acad Sci USA* 1998;**95**:12619−24.
44. Delport W, Cunningham M, Olivier B, Preisig O, van der Merwe SW. A population genetics pedigree perspective on the transmission of *Helicobacter pylori*. *Genetics* 2006;**174**:2107−18.
45. Raymond J, et al. Genetic and transmission analysis of *Helicobacter pylori* strains within a family. *Emerg Infect Dis* 2004;**10**:1816−21.
46. Kraft C, et al. Genomic changes during chronic *Helicobacter pylori* infection. *J Bacteriol* 2006;**188**:249−54.
47. Kraft C, Suerbaum S. Mutation and recombination in *Helicobacter pylori*: mechanisms and role in generating strain diversity. *Int J Med Microbiol* 2005;**295**:299−305.
48. Kennemann L, et al. *Helicobacter pylori* genome evolution during human infection. *Proc Natl Acad Sci USA* 2011;**108**:5033−8.
49. Covacci A, Telford JL, Giudice GD, Parsonnet J, Rappuoli R. *Helicobacter pylori* virulence and genetic geography. *Science* 1999;**284**:1328−33.
50. Linz B, et al. An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* 2007;**445**:915−8.
51. Montano V, et al. Worldwide population structure, long-term demography, and local adaptation of *Helicobacter pylori*. *Genetics* 2015;**200**:947−63.
52. Maixner F, et al. The 5300-year-old *Helicobacter pylori* genome of the Iceman. *Science* 2016;**351**:162−5.
53. Gray RD, Drummond AJ, Greenhill SJ. Lnguage phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 2009;**323**:479−83.
54. Renfrew C. Where bacteria and languages concur. *Science* 2009;**323**:467−8.
55. Latifi-Navid S, et al. Ethnic and geographic differentiation of *Helicobacter pylori* within Iran. *PLoS One* 2010;**5**:e9645.
56. Tay C, et al. Population structure of *Helicobacter pylori* among ethnic groups in Malaysia: recent acquisition of the bacterium by the Malay population. *BMC Microbiol* 2009;**9**:126.
57. Devi SM, et al. Ancestral European roots of *Helicobacter pylori* in India. *BMC Genomics* 2007;**8**:184.
58. Suerbaum S, Achtman M. *Helicobacter pylori*: recombination, population structure and human migrations. *Int J Med Microbiol* 2004;**294**:133−9.
59. Blaser MJ. Who are we? Indigenous microbes and the ecology of human diseases. *EMBO Rep* 2006;**7**:956−60.
60. Deleted in review.
61. Deleted in review.

Downloaded From Vet-eBooks.com

62. Comas I, et al. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat Genet* 2013;**45**:1176−82.
63. Smith NH, Hewinson RG, Kremer K, Brosch R, Gordon SV. Myths and misconceptions: the origin and evolution of *Mycobacterium tuberculosis*. *Nat Rev Microbiol* 2009;**7**: 537−44.
64. Hirsh AE, Tsolaki AG, DeRiemer K, Feldman MW, Small PM. Stable association between strains of *Mycobacterium tuberculosis* and their human host populations. *Proc Natl Acad Sci USA* 2004;**101**:4871−6.
65. Smith NH, et al. The population structure of *Mycobacterium bovis* in Great Britain: clonal expansion. *Proc Natl Acad Sci USA* 2003;**100**:15271−5.
66. Supply P, et al. Linkage disequilibrium between minisatellite loci supports clonal evolution of *Mycobacterium tuberculosis* in a high tuberculosis incidence area. *Mol Microbiol* 2003;**47**:529−38.
67. Wirth T, et al. Origin, spread and demography of the *Mycobacterium tuberculosis* complex. *PLoS Pathog* 2008;**4**:e1000160.
68. Gagneux S, Small PM. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infect Dis* 2007;**7**:328−37.
69. Hershberg R, et al. High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS Biol* 2008;**6**:e311.
70. Achtman M. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu Rev Microbiol* 2008;**62**:53−70.
71. Dos Vultos T, et al. Evolution and diversity of clonal bacteria: the paradigm of *Mycobacterium tuberculosis*. *PLoS One* 2008;**3**:e1538.
72. Pepperell C, et al. Bacterial genetic signatures of human social phenomena among *M. tuberculosis* from an aboriginal Canadian population. *Mol Biol Evol* 2010;**27**: 427−40.
73. Dou H-Y, et al. Associations of *Mycobacterium tuberculosis* genotypes with different ethnic and migratory populations in Taiwan. *Infect Genet Evol* 2008;**8**:323−30.
74. Mokrousov I. Genetic geography of *Mycobacterium tuberculosis* Beijing genotype: a multifacet mirror of human history? *Infect Genet Evol* 2008;**8**:777−85.
75. Brosch R, et al. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci USA* 2002;**99**:3684−9.
76. Mokrousov I, et al. Origin and primary dispersal of the *Mycobacterium tuberculosis* Beijing genotype: clues from human phylogeography. *Genome Res* 2005;**15**:1357−64.
77. Merker M, et al. Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nat Genet* 2015;**47**:242−9.
78. Gagneux S, et al. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* 2006;**103**:2869−73.
79. Monot M, et al. On the origin of leprosy. *Science* 2005;**308**:1040−2.
80. Monot M, et al. Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. *Nat Genet* 2009;**41**:1282−9.
81. Cole ST, et al. Massive gene decay in the leprosy bacillus. *Nature* 2001;**409**:1007−11.
82. Cardona-Castro N, et al. Human genetic ancestral composition correlates with the origin of *Mycobacterium leprae* strains in a leprosy endemic population. *PLoS Negl Trop Dis* 2015; **9**:e0004045.
83. Kolb AW, Ane C, Brandt CR. Using HSV-1 genome phylogenetics to track past human migrations. *PLoS One* 2013;**8**:e76267.
84. Holmes EC. Evolutionary history and phylogeography of human viruses. *Annu Rev Microbiol* 2008;**62**:307−28.

Downloaded From Vet-eBooks.com

85. Weber T, Major EO. Progressive multifocal leukoencephalopathy: molecular biology, pathogenesis and clinical impact. *Intervirology* 1997;**40**:98−111.
86. Chesters PM, Heritage J, McCance DJ. Persistence of DNA sequences of BK virus and JC virus in normal human tissues and in diseased tissues. *J Infect Dis* 1983;**147**:676−84.
87. Padgett BL, Walker DL. Prevalence of antibodies in human sera against JC virus, an isolate from a case of progressive multifocal leukoencephalopathy. *J Infect Dis* 1973;**127**: 467−70.
88. Kato A, et al. Lack of evidence for the transmission of JC polyomavirus between human populations. *Arch Virol* 1997;**142**:875−82.
89. Kunitake T, et al. Parent-to-child transmission is relatively common in the spread of the human polyomavirus JC virus. *J Clin Microbiol* 1995;**33**:1448−51.
90. Suzuki M, et al. Asian genotypes of JC virus in Japanese-Americans suggest familial transmission. *J Virol* 2002;**76**:10074−8.
91. Kitamura T, et al. Persistent JC virus (JCV) infection is demonstrated by continuous shedding of the same JCV strains. *J Clin Microbiol* 1997;**35**:1255−7.
92. Agostini HT, Yanagihara R, Davis V, Ryschkewitsch CF, Stoner GL. Asian genotypes of JC virus in Native Americans and in a Pacific Island population: markers of viral evolution and human migration. *Proc Natl Acad Sci USA* 1997;**94**:14542−6.
93. Sugimoto C, et al. Typing of urinary JC virus DNA offers a novel means of tracing human migrations. *Proc Natl Acad Sci USA* 1997;**94**:9191−6.
94. Jobes DV, Chima SC, Ryschkewitsch CF, Stoner GL. Phylogenetic analysis of 22 complete genomes of the human polyomavirus JC virus. *J General Virol* 1998;**79**: 2491−8.
95. Agostini HT, et al. Genotypes of JC virus in East, Central and Southwest Europe. *J Gen Virol* 2001;**82**:1221−331.
96. Sugimoto C, et al. Evolution of human polyomavirus JC: implications for the population history of humans. *J Mol Evol* 2002;**54**:285−97.
97. Agostini HT, et al. JC virus Type 2: definition of subtypes based on DNA sequence analysis of ten complete genomes. *J Gen Virol* 1998;**79**:1143−51.
98. Agostini HT, Ryschkewitsch CF, Brubaker GR, Shao J, Stoner GL. Five complete genomes of JC virus type 3 from Africans and African Americans. *Arch Virol* 1997;**142**: 637−55.
99. Guo J, et al. Geographical distribution of the human polyomavirus JC virus types A and B and isolation of a new type from Ghana. *J Gen Virol* 1996;**77**:919−27.
100. Stoner GL, et al. JC virus as a marker of human migration to the Americas. *Microbes Infect* 2000;**2**:1905−11.
101. Jobes DV, et al. New JC virus (JCV) genotypes from Papua New Guinea and Micronesia (Type 8 and Type 2E) and evolutionary analysis of 32 complete JCV genomes. *Arch Virol* 2001;**146**:2097−113.
102. Yanagihara R, et al. JC virus genotypes in the western Pacific suggest Asian mainland relationships and virus association with early population movements. *Hum Biol* 2002;**74**: 473−88.
103. Cui X, et al. Chinese strains (Type 7) of JC virus are Afro-Asiatic in origin but are phylogenetically distinct from the Mongolian and Indian strains (Type 2D) and the Korean and Japanese strains (Type 2A). *J Mol Evol* 2004;**58**:568−83.
104. Hatwell JN, Sharp PM. Evolution of human polyomavirus JC. *J Gen Virol* 2000;**81**: 1191−200.

Downloaded From Vet-eBooks.com

105. Chima SC, Ryschkewitsch CF, Fan KJ, Stoner GL. Polyomavirus JC genotypes in an urban United States population reflect the history of African origin and genetic admixture in modern African Americans. *Hum Biol* 2000;**72**:837−50.
106. Pavesi A. African origin of polyomavirus JC and implications for prehistoric human migrations. *J Mol Evol* 2003;**56**:564−72.
107. Pavesi A. Detecting traces of prehistoric human migrations by geographic synthetic maps of polyomavirus JC. *J Mol Evol* 2004;**58**:304−13.
108. Pavesi A. Utility of JC polyomavirus in tracing the pattern of human migrations dating to prehistoric times. *J Gen Virol* 2005;**86**:1315−26.
109. Wooding S. Do human and JC virus genes show evidence of host-parasite codemography? *Infect Genet Evol* 2001;**1**:3−12.
110. Fernandez-Cobo M, Agostini HT, Britez G, Ryschkewitsch CF, Stoner GL. Strains of JC virus in Amerind-speakers of north America (salish) and south America (Guaraní), Na-Dene-speakers of New Mexico (Navajo), and modern Japanese suggest links through an ancestral Asian population. *Am J Phys Anthropol* 2002;**118**:154−68.
111. Kitchen A, Miyamoto MM, Mulligan CJ. Utility of DNA viruses for studying human host history: case study of JC virus. *Mol Phylogenet Evol* 2008;**46**:673−82.
112. Shackelton LA, Rambaut A, Pybus OG, Holmes EC. JC virus evolution and its association with human populations. *J Virol* 2006;**80**:9928−33.
113. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 2005;**22**:1185−92.
114. Henne K, et al. Global analysis of saliva as a source of bacterial genes for insights into human population structure and migration studies. *BMC Evol Biol* 2014;**14**:190.
115. Dominguez-Bello MG, Blaser MJ. The Human microbiota as a marker for migrations of individuals and populations. *Annu Rev Anthropol* 2011;**40**:451−74.

Downloaded From Vet-eBooks.com

This page intentionally left blank

Downloaded From Vet-eBooks.com

Phylogenetic Analysis of Pathogens

8

D.A. Morrison
Uppsala University, Uppsala, Sweden

1. Introduction

Since the time of Charles Darwin, it has been considered important to be able to reconstruct phylogenies (the branching sequences of the lineages during their evolutionary history), both for a group of species and also for the individuals within those species. Unfortunately, this is one of the hardest forms of data analysis known. The events under study are unobservable historical events, and so we can neither make direct observations of them nor perform experiments to investigate them. Nevertheless, phylogenetics is based on the use of observable characteristics of contemporary organisms to try to deduce the sequence of events that occurred during the descent of those organisms. That is, we use what we can see now to infer the events that led to what we can see.

Charles Darwin's main contribution to biology was to recognize that there are two distinct types of biological evolution: (1) transformational evolution, in which individual objects each change through time and (2) variational evolution, in which groups of variable objects change their relative proportions through time. Transformational evolution is common in the physical sciences as well as in biology (e.g., the ontogeny of individuals). However, variational evolution has a special place in the biological sciences, because isolated changes in variation will ultimately lead to new species. Both types of evolution can best be represented as a tree- or network-like diagram (Fig. 8.1), because this can show the phylesis (changes through time via inheritance) as relative

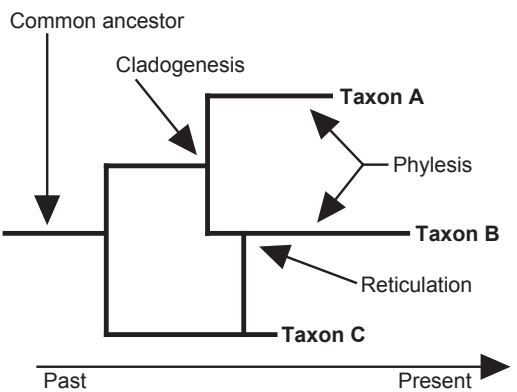


Figure 8.1 Phylogenetic tree for three contemporary taxa (A–C), showing the various relevant characteristics used for interpretation of the diagram.

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00008-1
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

branch lengths, the cladogenesis (speciation) as the relative branching order, and gene flow as reticulations among branches. The base of the tree/network represents the common ancestor, while the bifurcating and reticulating branches represent a successive series of descendants, arriving finally at the contemporary organisms at the twigs.

This chapter introduces the important facets of the topic of producing phylogenies, discussing those points that are of most direct relevance to the study of infectious organisms. I provide an overview, and an introduction to the recent literature. For more general issues, there are a number of essays,[1–4] including those specifically directed at prokaryotes,[5] as well as excellent books, varying from the introductory[6–9] to the detailed.[10–12]

2. The Uses of Phylogenies

Phylogenies form the framework within which we can best arrange our knowledge of all aspects of biology.[13,14] It has taken biologists a long time to explicate the simple idea that it is the study of biodiversity that makes biology different from other sciences—the nature and scale of the interrelationships among organisms is something that has never been conceived of within physics and chemistry. Evolutionary history is our explanation for the origin of that biodiversity, and so the best way to present biodiversity is in the context of phylogenetics. For pathogens, we are interested in the evolution of the diseases at the genetic level, and what this can tell us about their past and present diversity.

The term pathogen was devised about 1880, which was a time of great activity in the attempt to depict organismal relationships as trees, notably with the work of Ernst Haeckel. If we consider pathogenic organisms such as viruses, bacteria, microfungi, protists, and helminths, then it is clear that the members of some of these groups are not closely related to each other in the evolutionary sense, notably the organisms traditionally grouped as protists and helminths. Recognizing and understanding that these are utilitarian groupings based on nonevolutionary criteria has been one of the major contributions of phylogenetic analyses to modern biology.

Pathogens are often grouped on the basis of phenotypic similarity (e.g., hosts, predilection sites, infection route, and microscopy) or similarity of disease (e.g., symptoms and diagnostic procedures). However, these criteria do not automatically imply similarity of evolutionary history. For example, the traditionally recognized group helminths ("worms") consists of the Platyhelminthes (flatworms) and the Nematoda (roundworms). However, the latter have a body cavity enclosed by mesoderm (called the coelom) whereas the former do not, and so we infer that the roundworms are evolutionarily more closely related to (for example) insects than they are to the flatworms.

Perhaps the most valuable uses of a phylogeny are for both explaining and predicting organismal features. The strongest argument for a phylogenetic classification scheme is that it organizes our knowledge in a way that maximizes information content by being both explanatory and predictive.[15] For example, *Cryptosporidium* (Apicomplexa) causes cryptosporidiosis in mammals, and its life cycle and ultrastructure are

Downloaded From Vet-eBooks.com

similar to those of the agents causing coccidiosis, toxoplasmosis, and neosporosis in vertebrates; and thus it has traditionally been placed within the Coccidia. Molecular-based phylogenies contradict this placement,[16] however, indicating instead that it is related to the Gregarina, which infect invertebrates. This revised placement helps explain why anti-coccidial treatments are ineffective on members of *Cryptosporidium*: if susceptibility to anti-coccidial agents is a trait inherited from the common ancestor of the Coccidia, then any unrelated organisms will lack this trait.

Similarly, *Sarcocystis* is also part of the Coccidia, causing sarcocystosis in vertebrates. It has a two-host (indirect) life cycle, with the definitive host a carnivore and the intermediate host usually a herbivore. Sometimes, species have been collected only from the intermediate host, and we thus need to predict the definitive host. Our best prediction will be that the host is the same as for the closely related species. For example, *Sarcocystis alces* was originally collected only from European elk, the intermediate host, but Dahlgren and Gjerde[17] suggested that it is part of an evolutionary group that has canids as their definitive host; and so this would be our best prediction. This hypothesis was later tested successfully by Dahlgren and Gjerde,[18] who demonstrated that both red foxes and arctic foxes (canids) can act as definitive hosts.

There are many other important uses of phylogenies,[19] including the study of cophylogeny of hosts and pathogens (e.g., understanding the role of hosts in pathogen evolution) and pathogen biogeography (e.g., understanding the spread of pathogens). Different pathogens have different distributions, different patterns of spread, and different rates of evolution. This results in very different characteristics at the genetic and geographic levels. For example, the phylogeny can be compared to the geographic locations of the samples in order to investigate the spread of disease; or molecular dating methods[20] can be applied to estimate the age of important events in the origin and spread of new pathogens. These relationships are discussed in more detail in other chapters of this book.

3. The Logic of Phylogeny Reconstruction

Reconstructing a phylogenetic history is conceptually straightforward, although it took a long time for someone[21] to explicate the most appropriate approach. Interestingly, the study of historical linguistics has developed the same methodology,[22,23] thus independently arriving at exactly the same solution to what is, in effect, the same problem. From this point of view, the methodology itself is uncontroversial, and its generic nature means that it can be used for any objects with characteristics that can be identified and measured, and that follow a history of descent with modification.

The objective is to infer the ancestors of the contemporary organisms, and the ancestors of those ancestors, and so on, all the way back to the most recent common ancestor of the group of organisms being studied. Ancestors can be inferred because the descendant organisms share unique characteristics. That is, the descendants have features that they hold in common and that are not possessed by any other organisms. The simplest explanation for this observation is that the features are shared because

Downloaded From Vet-eBooks.com

they were inherited from an ancestor. The ancestor acquired a set of inheritable (i.e., genetically controlled) characteristics, and passed those characteristics on to its offspring. We observe the offspring, and from the resulting observations we infer the existence of the unobserved ancestor(s).

For example, we might note that a subset of all our organisms has an internal (bony) skeleton. No other organisms are known to possess this complex structure. There are only two realistic explanations for this observation: the organisms developed this structure independently, or they inherited it from a their common ancestor. The second explanation is the simplest one, and so it constitutes our working hypothesis of the evolutionary history of the organisms.

If we collect a number of such observations, what we often find is that they form a set of nested groupings of the organisms. For example, one subset of the organisms with an internal skeleton also possesses feathers, thus leading us to infer that this subgroup has a more recent common ancestor than does the skeleton group.

These nested sets and subsets of organisms can be represented in a tree diagram (Fig. 8.1), which has been the conventional way to denote hypotheses of phylogenetic history since the work of Charles Darwin. Each internal branch on such a tree indicates an inferred ancestor, and each terminal branch (or leaf) represents an observed organism. The branching order of the tree indicates the order of the historical events leading to divergence of the organisms, often called the "sister-group" relationships of the organisms. The length of the branches is commonly (but not always) used as a convention to represent the amount of evolutionary change that occurred in each ancestor, so that the length of a particular branch is proportional to the number of unique characteristics inferred to have been acquired by that ancestor (and passed on to its offspring).

These hypotheses of ancestry (both branching order and relative branch lengths) are open to testing by acquiring observations of other features of the organisms. These may support the previous observations or they may conflict with them. The practical process of reconstructing the phylogenetic history of a group of organisms consists of evaluating the (often) contradictory nature of the evidence. We collect as many observations as is practicable (given time, money, and other resources), and we compare the various pieces of evidence in order to arrive at the most plausible scenario for the historical events.

As a specific example that this logic can work in practice, Lemey et al.[24] studied the transmission history of the HIV-1 virus among a particular group of people. In this case, there was independent evidence concerning the transmission history, based on interviews with the nine people concerned, so that we have a pretty good idea who passed the virus to whom, and when. This known transmission history constitutes the true evolutionary history (Fig. 8.2). Some of the genes of the virus were also sequenced in these same people at varying time intervals. This means that we can independently attempt to reconstruct the evolutionary history (phylogeny) using these sequence data. In this case, the known history and the reconstructed phylogeny turn out to be identical, for at least some of the known types of phylogenetic analysis, and so we can justifiably conclude that our phylogenetic methods are valid.

As an example of an experimentally produced evolutionary history, we can consider the work of Sanson et al.[25] These workers used known errors in gene copying within *Trypanosoma cruzi* (Kinetoplastida), to mutate a single rRNA gene sequence

Downloaded From Vet-eBooks.com

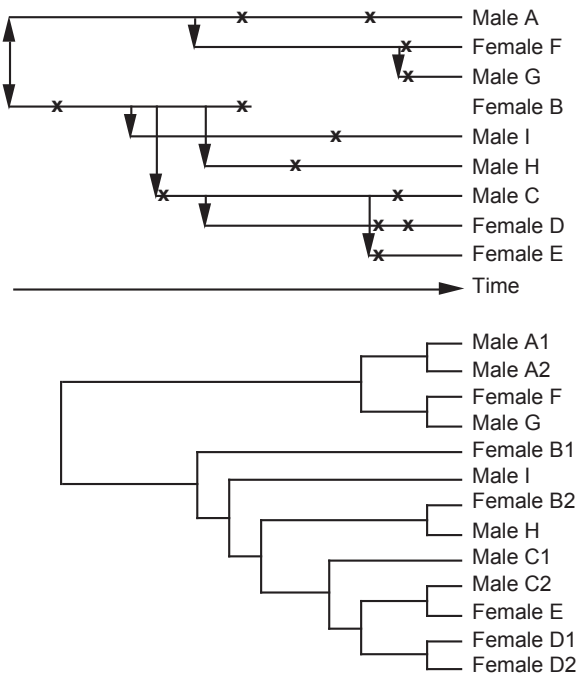


Figure 8.2 The known transmission history (upper panel) along with the phylogeny (lower panel) reconstructed from the *env* gene of the HIV-1 virus, based on the data of Lemey et al.[24] In the transmission history the *arrows* indicate the direction and time of transmission. So, for example, male A infected female F, who subsequently infected male G. The times at which samples were taken for DNA analysis are indicated by x. In the reconstructed evolutionary history, different samples from the same individual are indicated by labels 1 and 2. The inferred phylogeny is identical to the transmission history. For example, individuals F and G have a very closely related form of the virus, which is also closely related to the isolates A1 and A2.

into a set of eight descendant sequence clones, where all of the intermediate ancestral clones were sequenced as well. In this case, we have molecular data for all of the ancestors and all of the descendants, and we also know the true historical relationship among them all. Here, *all* of the known methods for reconstructing phylogenetic trees from molecular data produce exactly the same solution, which perfectly matches the known history.

4. Characters and Samples

Phylogenetic analysis can be used for any objects with characteristics that can be identified and measured, and that follow a history of descent with modification. The objects being sampled are usually referred to as "taxa" and the characteristics being measured are "characters."

The taxa can be any part of the standard taxonomic hierarchy (species, genera, families, etc.) or they can be individuals or populations. They can also be cultures or pathology samples, or even fossils. It is expected that the characters will be measurable on most of these taxa, although some of the taxa may lack some of the characters.

The sample of taxa used to construct the phylogeny needs to be adequate in order to provide a convincing case for particular phylogenetic relationships. Showing that a problem exists is easy with a small sample size, but revealing the solution usually takes

Downloaded From Vet-eBooks.com

much more effort. Furthermore, a biased sample usually leads to biased estimates of phylogenetic relationship. Relationships cannot be detected if the related groups have not been sampled, for example.

Unfortunately, pathogens are difficult to study because they are usually hard to find. Access to hosts can be difficult, and endoparasites often can be found only in symptomatic individuals. Therefore, sampling to date for many taxonomic groups has been almost entirely opportunistic.[26] Opportunities for sampling arise principally from case studies of medical and veterinary diseases, rather than from purposive experimental designs. Phylogenetic relevance has often not been the criterion for sample choice, which leads to a small and biased sample.

As a result of the large biodiversity of many pathogen groups, we need to choose exemplar taxa for a worthwhile phylogenetic analysis (e.g., at least one species from each genus). This is difficult if the biodiversity has not been well cataloged. In particular, the Apicomplexa, bacteria, and viruses are the three worst-known groups in terms of their named biodiversity, each with $<1\%$ of their estimated number of species having been described to date.[27] This compares very unfavorably with all other taxonomic groups. Even the Insecta, which is usually considered to be the prime example of a poorly known group, has about 1 million species known out of an estimated total of 4.5−30 million. This situation creates several possible impediments for the phylogenetic study of pathogens, which are discussed by Morrison[28] for the Apicomplexa as an empirical example.

Obviously, the characters measured must be heritable, which means that they must be genotypic characteristics rather than merely phenotypic ones (i.e., those greatly influenced by environmental variables). Most pathogens are unicellular or multicellular without specialized tissues, which severely limits the number and range of available characters. Traditionally, the characters used for phylogenetic and taxonomic analyses have been based mainly on life cycle features, disease characteristics, and ultrastructure. It may be rather difficult to determine homologies among such characters (i.e., their evolutionary comparability), so that related character states are being compared; and the data are also regrettably incomplete for most species. Consequently, phylogenies based solely on these characters have been rare, and they have not been particularly robust (see Barta[29] for an example).

For this reason, molecular data have now become the predominant character data for phylogenetic studies of pathogens. DNA mutates, the sequences change, and as pathogens spread they bring these changes with them. Molecular characters include allozymes, DNA−DNA hybridization, randomly amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), protein sequences, and nucleotide sequences. Of these, nucleotide sequences are now by far the most common, for all taxa not just pathogens. Indeed, many microbiology journals have guidelines stating that a phylogenetic context is required for the publication of new taxa, so that their nucleotide sequences and organismal phylogenies are part of the "publication pipeline."

For nucleotide sequences, only concordance between the phylogenies derived from several molecular sequences will be accepted as evidence for the organism phylogeny. A tree from a single molecular sequence represents only the phylogeny of that one

Downloaded From Vet-eBooks.com

gene, which does not necessarily reflect the phylogeny of the organism.[30] Just how many genes might be required to reconstruct the organismal phylogeny is an open question,[31,32] however. Many other molecular data types (e.g., AFLPs) sample widely from the genome, and so they should naturally represent the organismal phylogeny.

To date, most pathogen phylogenies have been based on the sequence of a single gene, usually the nucleotide sequence of the small-subunit (16S or 18S) ribosomal RNA gene. Indeed, most of the reclassification of the bacteria since the late 1970s has been based principally on this gene.[33] Many of the other genes sequenced are those for host recognition or for dealing with the host immune system (perhaps sequenced as part of projects producing new drugs or vaccines), which are often unique to each taxonomic group or are subject to heavy pressure from selection, and are thus not necessarily useful for phylogeny. In particular, bacterial genomes often have clusters of functionally related genes such as those for antibiotic resistance,[34] which can affect phylogenetic analysis. Consequently, the character data are rather fragmentary for many taxonomic groups. A multigene phylogeny is, therefore, unlikely to be produced from these current data (see Ogedengbe et al.[35] for an empirical example).

An obvious source of multigene sequence data is complete genomes, the necessary laboratory and data-analysis techniques being routinely feasible nowadays.[36,37] Thus, there are now available hundreds of complete genomes for bacterial taxa, although less than a score exist for the eukaryotic Apicomplexa, for example.

Sequences of complete genomes have contributed much to comparative genomics, which assumes that the phylogeny is known and can be used as the basis for comparisons among species. However, these genomes might never prove to be useful for phylogeny reconstruction itself. For example, there is likely to be increased homoplasy (owing to sequencing errors, intragene processes, intergene processes, and noncoding regions), along with inadequacies due to data-processing methods.[38]

The only phylogenetic situation where genome data are likely to be useful is where the original gene samples were biased, because the genomes might then correct the sampling error. However, if the genes previously examined were a representative sample of the genome, then the complete genomes will only confirm what was already known in terms of both confident and problematic relationships (e.g., see Morrison[38]). Of particular importance here is the possibility of horizontal gene flow (as opposed to the vertical inheritance often assumed by phylogenetics), which will be discussed in the following section.

We therefore need to be realistic about what we can expect from the phylogenetic analysis of sequence data, especially genomic sequences. Of particular importance will be our ability to locate representative genes that are appropriate to the evolutionary timescale being examined, rather than merely the quantity of the data per se. There needs to be a widespread base of people actively collecting a purposive sample of phylogenetically relevant multigene data.[39] Without this base, both the taxon and character sampling will be inadequate, in the sense that data will not be available for the critical exemplar taxa. This leads to uncertainties about organismal relationships, and concordance of multiple gene sequences cannot be demonstrated.

Downloaded From Vet-eBooks.com

5. The Practice of Phylogeny Reconstruction

Even though we cannot examine evolutionary history directly in any experimental way, scientists have developed sophisticated methods that allow us to use contemporary data about genotypic characteristics, to reconstruct phylogenies. While the logic of phylogeny reconstruction is straightforward, applying this logic in practice, in the face of conflicting evidence, is far from straightforward.

Phylogenetic analysis of molecular sequences (the predominant type of data these days) usually consists of three distinct procedures: (1) sequence alignment; (2) character coding; and (3) tree/network building. These steps are usually performed in this order, and all three of them need to be fully described for a phylogenetic analysis to be repeatable (in the scientific sense). Also, there are many known artifacts potentially associated with each of these procedures, and they need to be seriously considered in all analyses. Some of these issues and ways of dealing with them are illustrated by Morrison[3,27] using pathogens as examples.

Alignment is the process of establishing the possible homology relationships among the sequence residues.[39–43] Homology refers to the relationships of features that are shared among taxa due to common ancestry (i.e., they all inherited the feature from their most recent common ancestor). That is, we hypothesize that each of the aligned residues has descended from a common ancestral residue. Unfortunately, the term "homology" has been used historically to refer to a wide variety of concepts, and it is important to understand its strict evolutionary definition.[43,44]

Sequence similarity is often used as strong evidence for potential homology, and this is the basis of all automated alignment procedures (i.e., computer programs). However, sequence similarity decreases rapidly as taxa become more distant (in evolutionary time), so that processes causing sequence length variation become more probable (such as duplication, translocation, deletion, and insertion). Under these circumstances, similarity cannot be treated as homology (see Fig. 8.3). In evolutionary terms similarity = homology + analogy, and analogy (chance similarity due to character parallelism, convergence, or reversal), increases with increasing evolutionary distance. This exacerbates the problems of poor taxon sampling. It also exacerbates the problems caused by distant outgroups, which can be very difficult to align with the ingroup (see the following sections).

For the degree of sequence similarity that commonly occurs in phylogenetic analyses, automated alignment methods have often proved to be inadequate. For this reason, more than three-quarters of phylogeneticists manually intervene in the alignment process,[45] either by manually adjusting the alignment output by the computer program or by producing a completely manual alignment. This reflects the simple fact that there is not yet any automated procedure capable of producing a multiple sequence alignment that reflects homology.[42,43] Personal judgment may not be perfect, but at least it can consciously be based on homology as a concept.

For molecular data types other than sequences, homology often refers to homology of the bands appearing on the gels, and thus to the primers used. For example, AFLP data are based on a set of randomly chosen primers, and we must hypothesize that

Downloaded From Vet-eBooks.com

Taxa		Characters	49	49	49	49	49	49	49	49	50	50	50	50	50	50	50	50	50	50	50	51	51	51	51
1	L arlenae		A	T	C	A	A	G	C	G	–	T	T	G	T	G	C	G	C	T	T	C	A
2	L beveridg		A	T	C	A	A	G	C	G	–	T	T	G	G	G	C	G	C	T	C	C	A
3	L desclers		A	T	C	A	A	G	C	G	–	T	T	G	T	G	C	G	C	T	T	C	A
4	L discover		A	T	C	A	A	G	C	G	–	T	T	G	T	C	G	C	T	C	C	A
5	L elongatu		A	T	C	A	A	G	T	C	–	T	T	G	T	G	C	G	C	T	C	C	A
6	L gaevskay		A	T	C	A	A	G	C	G	–	T	T	G	T	G	C	G	C	T	T	C	A
7	L rachion1		A	T	C	A	A	G	C	G	–	T	T	G	T	G	C	G	C	T	T	C	A
8	L rachion2		A	T	C	A	A	G	C	G	–	T	T	G	T	G	C	G	C	T	T	C	A
9	L sommervi		A	T	G	A	A	A	C	G	–	T	T	G	T	G	C	G	C	T	C	C	A
10	L zubchenk		A	T	C	A	A	G	C	G	–	T	T	G	T	G	C	G	C	T	C	T	A
11	Neolepidap		A	T	C	A	A	G	G	C	G	A	T	G	G	G	C	G	C	T	C	C	A
12	Prod pried		A	T	C	A	T	G	T	G	–	T	T	G	C	G	C	G	C	T	C	C	A
13	Profundive		A	T	C	A	A	G	G	C	G	A	T	G	G	G	C	G	C	T	C	C	A
14	Myzoxenus		A	T	C	A	A	G	T	G	–	T	T	G	C	G	C	A	C	T	C	T	G
15	Intusatriu		A	T	C	A	G	G	T	G	–	C	T	G	C	G	C	G	C	T	C	T	G
16	Postlepida		A	T	C	A	A	G	C	C	–	T	T	G	C	G	T	G	C	T	C	T	G
17	Prod keyam		A	T	C	A	T	G	T	G	–	T	T	G	C	G	C	G	C	T	C	C	A
18	Lepidapedo		A	T	C	A	A	G	T	G	–	T	T	G	C	G	C	G	C	T	C	C	A
19	Clavogalea		A	T	C	A	T	G	T	G	–	T	T	G	C	G	C	G	C	T	C	C	A
20	Echeneidoc		A	A	C	A	T	G	T	G	–	T	T	G	T	G	C	G	C	T	C	C	A
21	Hypocreadi		A	A	C	A	C	G	T	G	–	T	T	G	T	G	C	G	C	T	C	C	A
22	Neomultite		A	A	C	A	G	G	T	G	–	T	T	G	C	G	C	G	C	T	C	C	A
23	Neohypocre		A	A	C	A	T	G	T	G	–	T	T	G	C	G	C	G	C	T	C	C	A
24	Lobatocrea		A	A	C	A	T	G	T	G	–	T	T	G	C	G	C	G	C	T	C	C	A
25	Diploproct		A	A	C	A	T	G	T	G	–	T	T	G	C	G	C	G	C	T	C	C	A

Taxa		Characters	49	49	49	49	49	49	49	49	50	50	50	50	50	50	50	50	50	50	50	51	51	51	51
1	L arlenae		A	T	C	A	A	G	–	C	G	T	T	G	T	G	C	G	C	T	T	C	A
2	L beveridg		A	T	C	A	A	G	–	C	G	T	T	G	G	G	C	G	C	T	C	C	A
3	L desclers		A	T	C	A	A	G	–	C	G	T	T	G	T	G	C	G	C	T	T	C	A
4	L discover		A	T	C	A	A	G	–	C	G	T	T	G	T	C	G	C	T	C	C	A
5	L elongatu		A	T	C	A	A	G	–	T	C	T	T	G	T	G	C	G	C	T	C	C	A
6	L gaevskay		A	T	C	A	A	G	–	C	G	T	T	G	T	G	C	G	C	T	T	C	A
7	L rachion1		A	T	C	A	A	G	–	C	G	T	T	G	T	G	C	G	C	T	T	C	A
8	L rachion2		A	T	C	A	A	G	–	C	G	T	T	G	T	G	C	G	C	T	T	C	A
9	L sommervi		A	T	G	A	A	A	–	C	G	T	T	G	T	G	C	G	C	T	C	C	A
10	L zubchenk		A	T	C	A	A	G	–	C	G	T	T	G	T	G	C	G	C	T	C	T	A
11	Neolepidap		A	T	C	A	A	G	G	C	G	A	T	G	G	G	C	G	C	T	C	C	A
12	Prod pried		A	T	C	A	T	G	–	T	G	T	T	G	C	G	C	G	C	T	C	C	A
13	Profundive		A	T	C	A	A	G	G	C	G	A	T	G	G	G	C	G	C	T	C	C	A
14	Myzoxenus		A	T	C	A	A	G	–	T	G	T	T	G	C	G	C	A	C	T	C	T	G
15	Intusatriu		A	T	C	A	G	G	–	T	G	C	T	G	C	G	C	G	C	T	C	T	G
16	Postlepida		A	T	C	A	A	G	–	C	C	T	T	G	C	G	T	G	C	T	C	T	G
17	Prod keyam		A	T	C	A	T	G	–	T	G	T	T	G	C	G	C	G	C	T	C	C	A
18	Lepidapedo		A	T	C	A	A	G	–	T	G	T	T	G	C	G	C	G	C	T	C	C	A
19	Clavogalea		A	T	C	A	T	G	–	T	G	T	T	G	C	G	C	G	C	T	C	C	A
20	Echeneidoc		A	A	C	A	T	G	–	T	G	T	T	G	T	G	C	G	C	T	C	C	A
21	Hypocreadi		A	A	C	A	C	G	–	T	G	T	T	G	T	G	C	G	C	T	C	C	A
22	Neomultite		A	A	C	A	G	G	–	T	G	T	T	G	C	G	C	G	C	T	C	C	A
23	Neohypocre		A	A	C	A	T	G	–	T	G	T	T	G	C	G	C	G	C	T	C	C	A
24	Lobatocrea		A	A	C	A	T	G	–	T	G	T	T	G	C	G	C	G	C	T	C	C	A
25	Diploproct		A	A	C	A	T	G	–	T	G	T	T	G	C	G	C	G	C	T	C	C	A

Figure 8.3 Two alternative alignments of part of the nuclear large-subunit ribosomal RNA gene for the Lepocreadioidea (Platyhelminthes). Two of the sequences have an extra nucleotide in length compared to the others, and this length variation needs to be addressed. The alignment on the left is the similarity-based alignment produced by the ClustalX computer program, which simply puts the extra nucleotides as far to the right as possible based on the relative "gap weights." The alignment on the right is based directly on putative homologies, implying in this case that the extra nucleotides have been formed as tandem repeats (or microsatellites). Note that there are considerable differences in the phylogenetic information, concerning the relationships of the taxa "Neolepidap" and "Profundive" to the other taxa, in two of the three columns that differ between the two alignments. The data are from Bray RA, Waeschenbach A, Cribb TH, Weedall GD, Dyal P, Littlewood DTJ. The phylogeny of the Lepocreadioidea (Platyhelminthes, Digenea) inferred from nuclear and mitochondrial genes: implications for their systematics and evolution. *Acta Parasitol* 2009;**54**:310–29.

Downloaded From Vet-eBooks.com

bands in the same position on the gels are homologous (i.e., they represent amplification by the primer pair of the same genetic sequence). The inability to assess these hypotheses is sometimes listed as a major limitation of nonsequence character data.

Character coding[46,47] is often overlooked as an important step in sequence analyses. Those parts of the sequence alignment involving length variation (where there are so-called "gaps") are sometimes considered to be uncertainly aligned, and most computer programs treat gaps as missing data. Furthermore, some regions in the sequence alignment might be considered to be ambiguously aligned across the dataset, even if some subsets of the sequences have been aligned with certainty. These regions often are excluded from the phylogenetic analysis.[48] In both cases, phylogenetic information is lost (see Fig. 8.4), even though this can potentially be useful for phylogenetics.[49] This issue can be dealt with by coding the length-variable regions as a set of independent characters, which are then included in the phylogenetic analysis.[46]

A cautionary note is warranted here. When dealing with nonmolecular character data, it is usual to decide a priori which characters will be sampled and which ones will not be. However, when collecting molecular data this only applies to the choice of genes to be sequenced or to the primers to be used. It does not apply to the actual data collected. This means that the experimenter is able to choose either to include or exclude the observations at will *after* the data have been collected. This applies when we decide to exclude characters that cannot be aligned unambiguously, alignment positions that appear to be overly variable or saturated (such as third-codon positions), or even simply positions where gaps have been introduced into the alignment.[3]

One approach to data selection would be to have some sort of measurement of our confidence in the alignment columns, which could be taken into account when the phylogeny is constructed. One practical problem with this approach is that there has been a veritable cottage industry developing such measurements, which have not yet been comparatively assessed for their suitability. So, there are objective criteria for deleting regions of variable or ambiguous alignment in phylogenetic analyses,[50–52] but a posteriori data exclusion should be treated with caution, as it has the obvious potential to introduce bias as well as to alleviate it.

Building the tree or network is the third step of a phylogenetic analysis, and it simply displays the information obtained from the sequence alignment and coding steps as a branching diagram.[3] That is, conceptually all it should do is change the tabular data (the alignment) into a picture of the data (the phylogeny), all of the hard work having been done in the previous two steps. In practice, it is rarely this simple.

In particular, it is important to remember that any genealogy is a network of relationships, irrespective of whether it represents relationships among individuals (a pedigree) or groups (a phylogeny). This is true any time that there is gene flow among contemporaries, in addition to vertical inheritance through time.[53] For phylogenies, this is true for any species that is subject to processes such as hybridization, introgression, lateral gene transfer, or other forms of recombination, for example.

In spite of this, phylogenetic trees are far more common in the biological literature than are phylogenetic networks, so that the reticulate relationships are ignored whenever they occur. This can often work in practice, because trees are a subset of networks. That is, trees are networks without reticulation, so that a tree is a special case of a network. A tree may

Downloaded From Vet-eBooks.com

(A) *Alignment*

```
Caecincola     CAGCAATGAGTACGGTTATATTGACTTGGC
Siphodera      CAGCAATGAGTACGGTAATATTGACTTGGC
Paracrypto     CAGCAATGAGTACGGTAATGCTGACATGGC
Mitotrema      CAGCAATGAGTACGGTAATGCTGACATGGC
Schistorch     CAGCACTGAGTACGGTAATCTGGAAATGGC
Callohelmi     CAGCATTGAGTACGGT-TTATGGACATGGC
Homal_arma     CAGCATTGAGTACGGT---ATGGACATGGC
Homal_syna     CAGCATTGAGTACGGT---ATGGACATGGC
N_splende1     CAGCATTGAGTACGGT---ATGGACATGGC
N_splende2     CAGCATTGAGTACGGT---ATGGACATGGC
```

(B) *Coding 1 — standard*

```
Caecincola     CAGCAATGAGTACGGTTATATTGACTTGGC
Siphodera      CAGCAATGAGTACGGTAATATTGACTTGGC
Paracrypto     CAGCAATGAGTACGGTAATGCTGACATGGC
Mitotrema      CAGCAATGAGTACGGTAATGCTGACATGGC
Schistorch     CAGCACTGAGTACGGTAATCTGGAAATGGC
Callohelmi     CAGCATTGAGTACGGT?TTATGGACATGGC
Homal_arma     CAGCATTGAGTACGGT???ATGGACATGGC
Homal_syna     CAGCATTGAGTACGGT???ATGGACATGGC
N_splende1     CAGCATTGAGTACGGT???ATGGACATGGC
N_splende2     CAGCATTGAGTACGGT???ATGGACATGGC
```

(C) *Coding 2 — gaps deleted*

```
Caecincola     CAGCAATGAGTACGGT   ATTGACTTGGC
Siphodera      CAGCAATGAGTACGGT   ATTGACTTGGC
Paracrypto     CAGCAATGAGTACGGT   GCTGACATGGC
Mitotrema      CAGCAATGAGTACGGT   GCTGACATGGC
Schistorch     CAGCACTGAGTACGGT   CTGGAAATGGC
Callohelmi     CAGCATTGAGTACGGT   ATGGACATGGC
Homal_arma     CAGCATTGAGTACGGT   ATGGACATGGC
Homal_syna     CAGCATTGAGTACGGT   ATGGACATGGC
N_splende1     CAGCATTGAGTACGGT   ATGGACATGGC
N_splende2     CAGCATTGAGTACGGT   ATGGACATGGC
```

(D) *Coding 3 —indels informative*

```
Caecincola     CAGCAATGAGTACGGTTATATTGACTTGGC 00
Siphodera      CAGCAATGAGTACGGTAATATTGACTTGGC 00
Paracrypto     CAGCAATGAGTACGGTAATGCTGACATGGC 00
Mitotrema      CAGCAATGAGTACGGTAATGCTGACATGGC 00
Schistorch     CAGCACTGAGTACGGTAATCTGGAAATGGC 00
Callohelmi     CAGCATTGAGTACGGT?TTATGGACATGGC 10
Homal_arma     CAGCATTGAGTACGGT???ATGGACATGGC 11
Homal_syna     CAGCATTGAGTACGGT???ATGGACATGGC 11
N_splende1     CAGCATTGAGTACGGT???ATGGACATGGC 11
N_splende2     CAGCATTGAGTACGGT???ATGGACATGGC 11
```

Figure 8.4 Alignment of part of the nuclear large-subunit ribosomal RNA gene for the Lepocreadioidea (Platyhelminthes). The alignment (A) has several taxa with a gap that might be phylogenetically informative, and which can be coded in any of several ways that do not represent the same phylogenetic information (B−D). Most phylogeny programs treat the gaps as missing data (B), so that each alignment column independently contributes information only

Downloaded From Vet-eBooks.com

thus be a useful model in practice,[54] but we should not lose sight of the fact that a tree is actually a simplified network—all trees are networks but not all networks are trees.[38]

A number of different types of data analysis have been developed, based on different mathematical optimality criteria. Some of these are based on estimated genetic distances while others are based directly on the characters, such as parsimony, likelihood, and bayesian analysis. The latter try to maximize the amount of inferred homology in the phylogeny (or minimize the amount of inferred homoplasy) as part of their optimality criterion, which gives them a theoretical advantage (and one that also appears in practice). Choosing among such methods is discussed in the following section.

Unfortunately, different phylogeny-building methods often add artifactual information to the tree/network that does not reflect evolutionary history. For example, substitutional saturation is an almost universal problem (due to superimposed substitutions[55]) and compositional heterogeneity is a recurring problem (e.g., A + T bias or codon bias[56]), as are juxtaposed long and short branches (resulting in what is known as long-branch attraction[57]). It is worth noting that many of the currently recognized practical problems (e.g., long-branch attraction, compositional bias, and saturation) are merely specific examples of how analogy appears in molecular biology. Analogy exacerbates the problems caused by poor taxon sampling and distant outgroups.

While it is impossible to make generalizations about the phylogenetic problems of pathogens, because the different groups are not closely related, there are recurring themes. For example, the main cause of substitutional saturation and long-branch attraction is large evolutionary distances among the taxa, which is a common situation for unicellular organisms such as most pathogens. Similarly, nucleotide composition biases reflect mutational as well as selective forces, so that AT-richness often characterizes mutation-prone genomes such as those of intracellular bacteria, although there are also bacteria (such as the Actinobacteria) that are GC-rich instead. Nucleotide bias is also associated with the parasitic lifestyle, such as in the AT-richness of *Plasmodium falciparum* (Apicomplexa), where it is presumably advantageous because it permits rapid genetic selection in response to survival threats.

Computationally, artifacts arise because one or more of the assumptions of the analysis have been violated. All data analyses are based on some form of underlying model, whether explicit or implicit, which specifies the assumptions that need to be met by the data in order for the results of the analyses to be reliable.[58] The choice among phylogenetic models should be quantitatively assessed rather than arbitrarily chosen,[59] as this is the only proactive way of dealing with artifacts. These issues often

for those taxa with nucleotides in that column. Here, the gaps are not treated as indels, but as missing information. Alternatively, many researchers simply delete alignment columns that contain gaps (C), thus losing all of the potential phylogenetic information. Here, the indels do not exist at all. Other researchers code the gapped columns as separate indels (D). Here, extra characters are added that represent the sharing of the indel patterns among the taxa, which are then phylogenetically informative when analyzed.
The data are from Bray RA, Waeschenbach A, Cribb TH, Weedall GD, Dyal P, Littlewood DTJ. The phylogeny of the Lepocreadioidea (Platyhelminthes, Digenea) inferred from nuclear and mitochondrial genes: implications for their systematics and evolution. *Acta Parasitol* 2009;**54**: 310–29.

Downloaded From Vet-eBooks.com

can be dealt with by deleting length-variable regions and autapomorphies from the alignment, or by choosing appropriate evolutionary models for the analysis.[3]

The most basic assumption of the models is that the model does not change through time along the evolutionary lineages (i.e., in different subtrees). If this is so, then mathematically the model is said to be stationary. Biologically, stationarity is an unlikely assumption, because the physical constraints on the macromolecule coded for by the gene are likely to have varied through time, and so the DNA sequence is expected to also have been subjected to temporal variation. Suggestions have been made that allow for temporal variation in parameters of likelihood models.[3,60] Unfortunately, few of the current suggestions are yet to be incorporated into the most commonly used computer programs, mainly because they do not fit easily as extensions of the current simple models.

Phylogenetic analysis of all organisms is usually treated as being rather similar, except for viruses and perhaps bacteria. Otherwise, the differences between different pathogen groups are quantitative rather than qualitative. Some groups have certain genotypic characteristics more strongly than do others, and these will thus affect the analyses to varying degrees. Bacteria often are subject to horizontal gene flow of some sort, as well as hierarchical inheritance, and this can confound phylogenetic inferences. This is discussed in more detail in a later section. For viruses, it is often possible to study the genotypic changes occurring during the course of infection from serial samples, due to their rapid evolution. Suitable methods for the phylogenetic analysis of serial samples are currently under development.[61,62] This clearly has implications for genome-wide association studies.[63]

6. Choosing a Method

It is possible to perform all three procedures of a phylogenetic analysis (sequence alignment, character coding, and tree/network building) simply by choosing some popular computer programs and then using the default parameter values of those programs. For example, one strategy popular in the literature is to choose Clustal for alignment, to ignore any explicit coding, and then to choose MEGA for tree building.[64] Unfortunately, this is a very naïve approach, because it does not consider the possible unsuitability of the analyses for the specific dataset at hand, which may lead to results that are artifacts.[65] A phylogenetic analysis is only as good as the steps taken to ensure the highest quality of data and to evaluate and use the most appropriate mathematical models for the data analysis.

Unfortunately, in some areas of biology overly simplistic analyses still seem to be the order of the day for many practitioners. In the modern world, however, with the advent of more realistic models of character evolution, phylogenetic analyses need no longer be treated as "black boxes" into which data are fed and from which a tree spontaneously emerges. We need to be aware of what assumptions are made by different analyses and how to interpret the information that comes out. This knowledge will help to choose an appropriate phylogenetic analysis for the data.

This chapter is not the place to review the pros and cons of each and every method, and this can be found in several books.[10–12] You will find that there are several important concepts to bear in mind when considering different methodologies: efficiency; the objective function used; the search strategy used (exhaustive, branch-and-bound, and heuristic); robustness; power; consistency; reconstruction probability; and falsifiability. The

Downloaded From Vet-eBooks.com

method chosen will probably be a compromise from among these criteria, as no method has yet shown itself to be superior on more than a few of them.

There are two distinct types of error that will affect a phylogenetic analysis: (1) random or stochastic error and (2) systematic error or bias. Stochastic error is variation that results from sampling. That is, we cannot make a complete inventory of all of the data that could be collected, and so we collect a sample instead. That sample may not be representative of the complete collection of data, and this results in random error. Systematic error, on the other hand, results from mismatches between our goal and our sampling and analytical procedures. That is, we may (unintentionally) collect data from taxa that are inappropriate (e.g., diseased), or choose to analyze the data with an inappropriate evolutionary model. Systematic error is thus associated with the accuracy of the answer (i.e., how close to the truth we get), while random error is associated with the precision with which we can present that answer (i.e., how repeatable it is).

In a phylogenetic study, random error is always expected to occur, but we can attempt to reduce its impact, while systematic error is something that we actively strive to avoid if we can. Random error can usually be dealt with by increasing the sample size, either of characters or of taxa as appropriate.[32] Systematic error, however, cannot be fixed by increases in sample size because the same bias will exist throughout the genome.[66] For example, several of the gene trees of the Microsporidia have been shown to suffer problems with long-branch attraction due to fast-evolving lineages,[67] and this is not alleviated by studying whole genomes because these fast-evolving genes occur genome-wide. If systematic bias affects many or most of the genes then the reconstructed organismal phylogeny will be wrong, and adding new genes will not resolve the issue. Similar problems have been reported for whole genomes of the Apicomplexa, where incongruent phylogenetic relationships based on a small number of genes were simply confirmed as incongruent by whole-genome phylogenies.[38]

Of particular importance is the wide diversity of known biological processes that can obscure the genetic patterns produced by phylogenetic history. That is, our data display a set of phylogenetic patterns produced by a set of phylogenetic processes, and the aim of the data analysis is to reconstruct the process history from the observed patterns. This is a hard task because the same patterns can be produced by any of several processes.[68] This means that almost all datasets will show incongruent genetic patterns, from which we attempt to construct the species phylogeny. These incongruences will arise from the following sorts of processes[69]:

- intergenomic transfers (nuclear copies of mitochondrial DNA and nuclear copies of plastid DNA);
- horizontal gene flow (hybridization, introgression, horizontal gene transfer, and plastid capture);
- lineage sorting stochasticity (deep coalescence);
- genome organization (number of chromosomes, ploidy level, gene linkage, and gene duplication−loss);
- demography (effective population size);
- natural selection (bottlenecks and selective sweeps);
- phylogeographic structure (spatial arrangement of the genetic structure).

Downloaded From Vet-eBooks.com

Some of these will create tree-like phylogenies and some will not. The "art" of phylogeny reconstruction is to separate the patterns due to vertical inheritance and horizontal gene flow, if possible.

As the number of multigene datasets increases, an important methodological decision will therefore be how best to derive the organismal phylogeny from a collection of (usually incongruent) gene phylogenies (i.e., how to get the species phylogeny from the gene trees). Note that there are actually two separate issues here. First, a phylogeny produced from any one dataset may or may not represent the true history of the taxa in that dataset (e.g., the reconstructed gene tree might not be the true gene tree). Second, even if we have the true tree for the dataset it still may or may not represent the true history of the taxa (e.g., the gene tree might not be the same as the species tree). Indeed, there are compelling reasons to expect that most gene trees will not match the species phylogeny.[70] Dealing with both of these issues simultaneously is no mean task (reviewed by Nakhleh[71] and Szöllösi et al.[72]).

There are two basic strategies for analyzing combined data from multiple datasets[3]: (1) combine the data into one set and then produce a single phylogeny from it and (2) produce a tree from each of the datasets and then combine these into a single phylogeny. That is, we can do the combining either before or after we construct the tree/network. The first strategy can be called concatenation (since we concatenate the data) while the second can be called consensus (since we produce a consensus of the trees), although these strategies have been called many different things in the literature (e.g., supermatrix and supertree, respectively). These two strategies may produce mutually contradictory answers, although they often do not, and there is a long history of unresolved debate concerning their relative merits.[73–75] Indeed, methods are under constant development to improve upon these approaches by estimating the organismal phylogeny directly rather than indirectly.[76–78]

7. Representing Phylogenies: Trees

Almost all early representations of biological relationships involved networks, not trees; indeed, the tree icon was introduced explicitly as a simplification of a network.[79] It was Charles Darwin who popularized the idea of using a tree metaphor for genealogical relationships, emphasizing the so-called Tree of Life; and this iconography then came to dominate phylogenetics during the 20th century. This has especially been true with the development of quantitative phylogenetic methods, where an explicit mathematical model is required (rather than simply a visual metaphor).[38]

The idea of a tree as the appropriate representation of phylogenetic relationships has thus been with us for 150 years now, and yet it is quite clear from the literature that many biologists have still not fully grasped this idea and its consequences.[9] That is, misinterpretation of trees, and the taxon groupings (clades) represented by those trees, is endemic in comparative biology.[80–83] Indeed, this failure of "tree thinking" seems to be deep-seated in the general public, as well.[84,85]

An evolutionary tree obviously must have a time direction (from ancestors to descendants), which is provided by the root. That is, the internal nodes of the tree represent ancestors and the external nodes represent the final descendants. If the taxa were

Downloaded From Vet-eBooks.com

species, each node would then represent a speciation event and the branch lengths would represent the amount of change in the sequences.

An unrooted tree cannot be a picture of evolutionary history because there is no indication of the direction of evolutionary change across the tree (which would be away from the root). However, an unrooted tree can be an important step toward obtaining a picture of evolutionary history. For example, for nine taxa there are 135,135 unrooted binary trees each of which can be rooted in any one of 15 different places (Fig. 8.5), yielding 2,027,025 possible rooted trees. Finding the unrooted tree thus eliminates 2,027,010 of these trees, leaving us with only 15 possible trees. This is clearly a major step, even if we never work out the precise location of the root.

Nevertheless, almost all of the questions being asked by biologists, which they are trying to answer by performing a phylogenetic analysis, can only be answered using a rooted diagram. It is inappropriate to identify evolutionary "groups" of taxa on an unrooted tree,[3,86] because only monophyletic groups (called clades) make any sense in an evolutionary context. A clade includes the most recent common ancestor of the group plus all of its descendants; and so, by definition, a clade cannot be determined from an unrooted tree. An unrooted tree only indicates partitions (or splits) in the collection of taxa. For example, there are three possible ways to split four taxa into partitions of two taxa each, and an unrooted tree will show only one of them. Thus, an unrooted tree contains information that allows us to *eliminate* possible groups

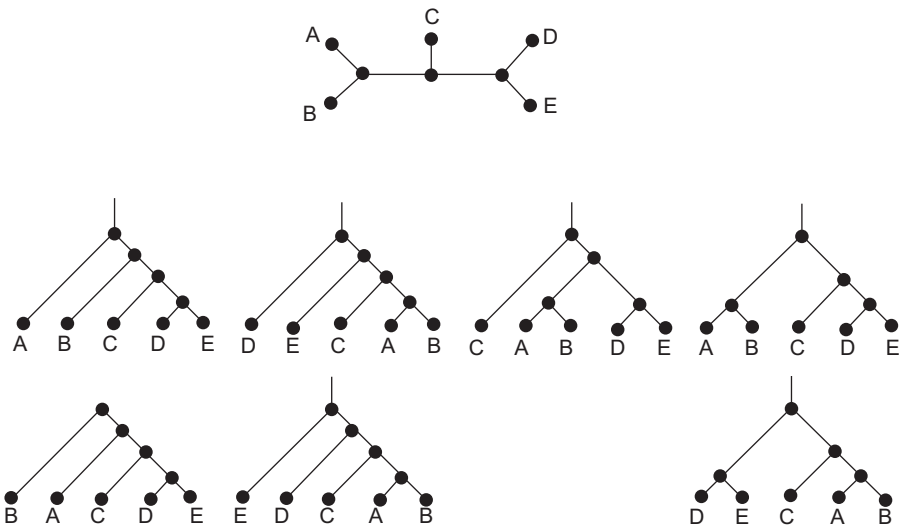


Figure 8.5 An unrooted tree of five taxa (top), which can be rooted on any of its seven branches, yielding seven rooted trees (below). It should be clear which branch of the unrooted tree has been used as the root for each of the seven rooted trees. Thus, there are more rooted trees than unrooted trees, because every unrooted tree can potentially be rooted on any of its branches.

Downloaded From Vet-eBooks.com

from consideration. However, it does not contain positive information about groups, because not necessarily both of the partitions form evolutionary groups.

Furthermore, relationships among clades are equal, in the sense that each clade is the sister to some other clade and vice versa. Thus, clades cannot be "basal" or "crown,"[83,87,88] because each single clade branches from some other single clade, rather than each clade being a side-branch from a main stem. Logically, at each speciation event two new species arise, rather than one species producing an extra offshoot species. There is no main stem in an evolutionary tree, but instead there is a series of branches leading to a series of twigs, even if some of the branches do have more twigs than others. Furthermore, neither of the sisters represents the ancestor; instead, they share a common ancestor, which may not look like either of them.

Finally, characters change through time, and so character states can be either ancestral (the original form) or derived (modified in some descendant). However, clades themselves cannot be either ancestral/primitive ("lower") or derived/advanced ("higher"), as each clade will have a combination of ancestral and derived character states.[87] There is no chain leading from ancestral species to derived species. Instead, each species (or group) is the sister to some other species (or group), with which it shares some characters inherited from their ancestors and from which it differs by some unique characters. Any group that is interpreted to be ancestral is paraphyletic (since it does not contain all of the descendants from the common ancestor) rather than monophyletic, and thus has no phylogenetic relevance.

All of this leads us inevitably to the question of how best to root a phylogenetic tree. For molecular data, there are basically six ways that have been proposed[3,89]: (1) a priori polarizing of the character states; (2) via reversible substitution models; (3) midpoint rooting; (4) using the molecular clock, or minimizing tip−root variance; (5) coalescence theory for population samples; and (6) using an outgroup. Some of these methods have been more popular than the others, and not all of them are equally effective.[90–92]

Use of an outgroup, (6), is far and away the most widespread method of rooting, and rightly so. The outgroup consists of one or (preferably) more taxa that are not part of the study group (i.e., the ingroup). The root of the tree is then simply the branch that connects the outgroup taxa to the ingroup taxa. The main limitation of this method is the choice of the taxa to be included in the outgroup. For robust phylogenetic analysis[93] the outgroup needs to consist of several members of the sister taxon to the ingroup (i.e., the most closely related group to the ingroup), preferably ones with relatively short branch lengths to the ingroup. Evolutionarily more-distant species can end up rooting the ingroup at what is effectively a random location, due to the lack of relevant phylogenetic signal involved in the long branch lengths leading to the outgroup. Alternatively, evolutionarily close species may not be reciprocally monophyletic with the ingroup, due to incomplete separation of their gene flows; this means that there will be multiple "true" root locations on the tree. So, choosing an appropriate outgroup is a balancing act between too close and too distant, even for genomic datasets.

The only way to root the Tree of Life, which is of some interest when dealing with pathogens, since many of these groups were intimately involved in the origin of life, is

Downloaded From Vet-eBooks.com

to use method (1). This has been a topic of long-standing interest in evolutionary biology.[93,94]

8. Phylogenetic Networks

The view of phylogenetics described in the previous section assumes a hierarchy of bifurcating (or sometimes multifurcating) groups. Indeed, the assumption of a universal Tree of Life hinges upon the process of evolution being tree-like throughout history.[13,14] In eukaryotes, the molecular mechanisms and species-level population genetics of variation mainly do cause a tree-like structure over time, except in groups where horizontal gene flow is common (e.g., plants and fish). However, in prokaryotes, these processes often do not lead to a tree-like structure, as there are known to be many mechanisms for genetic exchange that disrupt a genealogical tree.[95]

This has led to an ongoing discussion about whether bacterial phylogenetics, in particular, should be based on the concept of a tree[96] or not.[97] We have previously used a series to represent biodiversity (the Great Chain of Being) and we have used a tree (the Tree of Life)—does our increased understanding of molecular evolution mean that it is time to find a new representation[53]?

To this end there has been much interest in the use of networks rather than trees as the basis for phylogenetic analysis. The intention here is to replace the Darwinian model of a bifurcating tree by a "reticulating tree,"[98] with the reticulations representing evolutionary processes other than lineal descent with modification. Such processes involve gene flow of some sort, including: hybridization, introgression, recombination, horizontal (or lateral) gene transfer, genome fusion, ancestral polymorphism (also called deep coalescence or incomplete lineage-sorting), and gene duplication−loss (or hidden paralogy). The difficulties of fitting bacteria[99] and hybrids[100] into a phylogenetic tree were first aired before 1985, over 30 years ago, but the issues have only recently received widespread attention.

Unfortunately, this field is rather poorly developed at the moment.[101] Network methods that try explicitly to represent evolutionary history (called "evolutionary networks") all have serious restrictions on the types of patterns they can analyze, and on the allowed complexity of those patterns (see Morrison[102] for a 1997 review). As noted by Huson et al.,[103] there are many promising directions to follow and rudimentary software implementations, but there is no tool currently available that biologists can routinely use on real data.

All of the discussion in the previous section (about trees) applies equally to networks, because a tree is merely a special case of a network. An evolutionary network must thus be rooted, in order to form a hypothesized evolutionary history. All of the internal nodes should be (inferred) ancestors and all of the branches should represent inferred evolutionary events (with a direction of transformation). Nodes where two or more lineages converge (reticulation nodes) indicate pooling of genetic material; and nodes with one branch coming in and two or more going out (tree nodes) represent genetic divergence (see the empirical example in Fig. 8.6).

Downloaded From Vet-eBooks.com

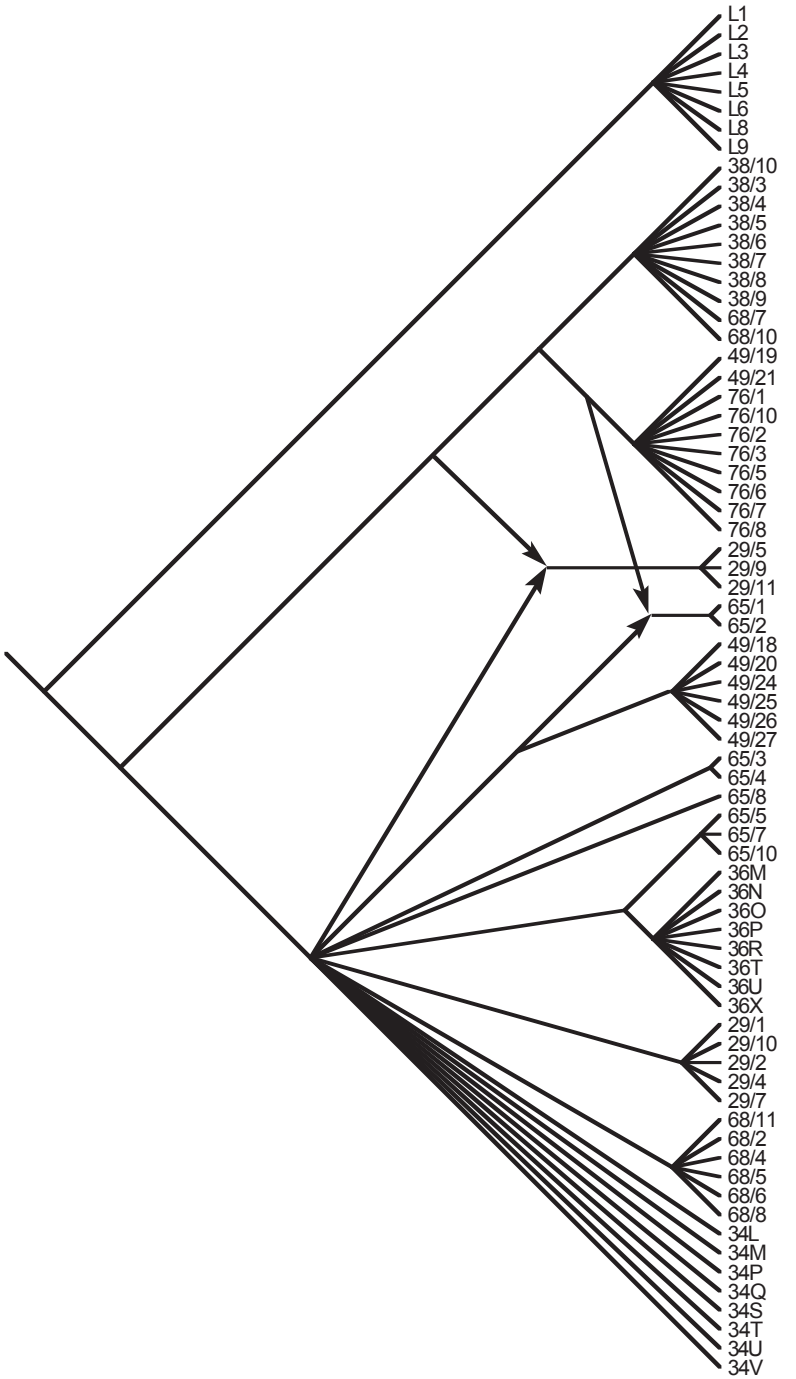


Figure 8.6 Recombination network for 1542 aligned nucleotides from 72 samples of *Dictyocaulus viviparus* (Nematoda). This evolutionary network shows the inferred historical relationships among 64 farm samples and 8 samples from a laboratory isolate (used to root the network, at the left). Most of the samples from each farm are closely related in a simple divergent fashion through time. However, two groups of samples descend from reticulation nodes (indicated by *arrows*), thus indicating the pooling of two distinct sources of genetic material. The farms involved (29 and 65) may thus have multiple sources of infection. The data are mitochondrial protein, rRNA, and tRNA gene sequences from Höglund et al.[110]

Downloaded From Vet-eBooks.com

There are, of course, an additional set of considerations for evolutionary networks compared to trees, given their increased complexity resulting from horizontal gene flow. While it is conceptually straightforward to generalize (to a network) the algorithmic approaches previously developed for trees, it is very difficult in practice. Essentially, the tree models are extended from their simple form, which includes only evolutionary processes such as nucleotide substitutions and insertions/deletions, to also include the reticulation events. We can conceptualize a reticulating network as a set of interlinked trees, and if we do so then the optimization procedure can be seen as optimizing one set of characteristics within each gene tree and optimizing another set of characteristics across the set of trees. The main practical difficulty then becomes the much greater mathematical space occupied by the solutions to the optimization problems.

However, not all networks are simply trees with extra edges (which is sometimes called an augmented tree or a tree-based network)—there exist networks that are inherently network-like and cannot be obtained by adding reticulation edges to a tree.[104] In these cases, the concept of an "underlying tree" is meaningless, and these networks will require new conceptual models as well as different mathematical models. This issue seems to be particularly pertinent to bacterial phylogenetics.

Given the current limitations, instead of evolutionary networks, what we have more commonly is a wide array of methods for displaying data conflict in phylogenetic datasets (called "data-display networks"). That is, compatible or congruent data patterns are displayed as a tree, while incongruences in the data are displayed as reticulations in the tree. An empirical example of constructing such a network is shown in Fig. 8.7.

Incongruences can also arise, in addition to the gene-flow processes listed earlier, from: (1) analogous rather than homologous characters (e.g., parallelism, convergence, and reversal); or (2) methodological issues in data collection (e.g., taxon sampling, character sampling, and outgroups) or data analysis (e.g., model mis-specification and choice of optimality criterion). We cannot distinguish, from the network alone, the cause of any character incongruences,[68] and so the nodes of the network do not necessarily represent ancestors (as they would in a rooted network), and the branches do not necessarily represent biological character transformations (from ancestor to descendant). Data-display networks are very useful for exploratory data analysis[105] or estimating genetic diversity,[106] but they should not be confused with (or treated as) evolutionary networks.

It is becoming increasingly obvious that the basic biological model for most evolutionary studies is a phylogeny that includes nontree (reticulation) events, especially when dealing with whole genomes. Since most gene trees are not expected to match the species phylogeny, even when it is tree-like, when is it worthwhile to reconstruct a species tree? Resolving this issue, and devising methods for constructing evolutionary networks, may be the biggest current challenges for phylogenetics, particularly for bacteria.[106–109]

Much of the problem arises from the lack of sexual reproduction and lack of available macro-characters in prokaryotes, so that molecular mechanisms loom large in their phylogenetic datasets, particularly horizontal gene transfer. Furthermore, sequences of the small-subunit rRNA gene have played the dominant role in microbiology, and one gene phylogeny cannot be used reliably to reconstruct the

Downloaded From Vet-eBooks.com

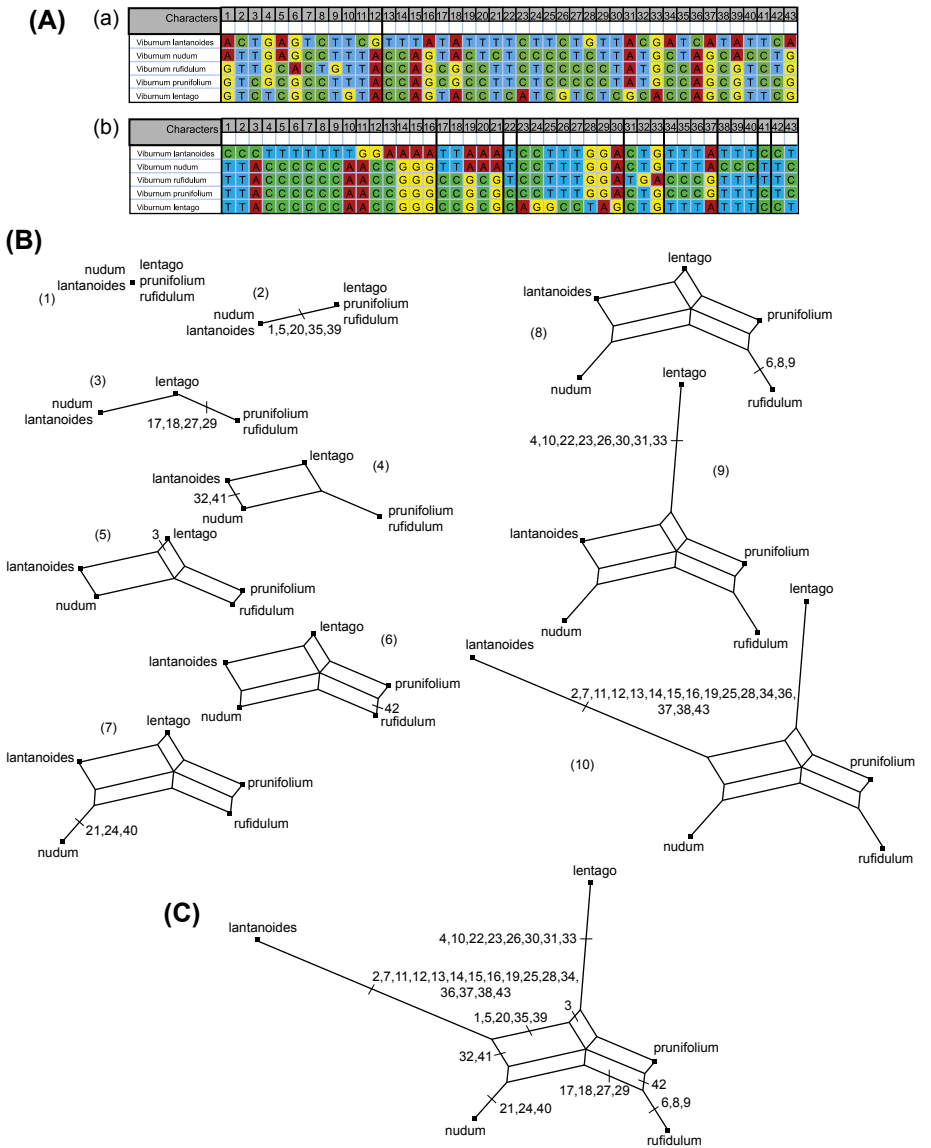


Figure 8.7 Construction and interpretation of a Median network of 43 variable nucleotides (out of 1687 sequenced) from five *Viburnum* species (Plantae). (A) The aligned nucleotide positions in their original sequence order (a), and rearranged to group identical patterns of variation across the species (b). This shows that the data support nine distinct patterns of relationship among the sequences out of the 15 patterns possible (i.e., there is considerable incongruence in the dataset). (B) The steps (1−10) required to construct the network from the aligned sequences. The species start in one group (step 1), and then each of the nine patterns of relationship is added to the network by sequentially creating subgroups of species (steps 2−10). At each step, an edge (or set of parallel edges) is added to the growing network, representing one of the data patterns. The characters involved at each step are marked on the appropriate network edge, numbered according to their original sequence order. The length of each edge is proportional to the number of characters defining (or supporting) that edge. (C) The completed network, with the supporting characters labeling each edge. The network displays all nine of the incongruent data patterns, which cannot be done using a tree. The data are chloroplast *trn*K intron and nuclear ribosomal ITS gene sequences from Donoghue et al.[111]

Downloaded From Vet-eBooks.com

organismal evolutionary history. The sequences of the small-subunit rRNA gene may well have a tree-like history but that does not automatically entail that the genomes have a similar structure.

The rest of the problem comes from whether we see the Tree of Life primarily as a metaphor (i.e., a model) for the structure of the evolutionary past, or whether it is a specific hypothesis about that structure (i.e., the evolutionary process really does generate a tree). Obviously, there is a tree-like history generated by cell divisions of prokaryotes, but is this "Tree of Cells" the most useful way of organizing our knowledge of biodiversity? Microbiologists seem to have been at times wary of phylogenetic analysis, and much of the history of bacterial classification has unfolded by deliberately ignoring the basic principles that I have summarized here.[33] Indeed, it may be that microbiology and phylogeny are incompossible. If so, then microbiologists need another paradigm; but those who object to trees do not yet seem to have one (that is, they are anti-tree rather than pro-something-else).

References

1. Stevens JR, Schofield CJ. Phylogenetics and sequence analysis: some problems for the unwary. *Trends Parasitol* 2003;**19**:582−8.
2. Delsuc F, Brinkmann H, Philippe H. Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet* 2005;**6**:361−75.
3. Morrison DA. Phylogenetic analyses of parasites in the new millennium. *Adv Parasitol* 2006;**63**:1−124.
4. Yang Z, Rannala B. Molecular phylogenetics: principles and practice. *Nat Rev Genet* 2012;**13**:303−14.
5. Williams TA, Heaps SE. An introduction to phylogenetics and the tree of life. In: Goodfellow M, Sutcliffe I, Chun J, editors. *New approaches to prokaryotic systematics*. London: Academic Press; 2014. p. 13−44.
6. Bromham L. *Reading the story in DNA: a beginner's guide to molecular evolution*. Oxford: Oxford University Press; 2008.
7. Lemey P, Salemi M, Vandamme A-M, editors. *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*. 2nd ed. Cambridge: Cambridge University Press; 2009.
8. Hall BG. *Phylogenetic trees made easy: a how-to manual*. 4th ed. Sunderland (MA): Sinauer Associates; 2011.
9. Baum DA, Smith SD. *Tree thinking: an introduction to phylogenetic biology*. Greenwood Village CO: Roberts and Co.; 2012.
10. Nei M, Kumar S. *Molecular evolution and phylogenetics*. New York: Oxford University Press; 2000.
11. Felsenstein J. *Inferring phylogenies*. Sunderland (MA): Sinauer Associates; 2004.
12. Yang Z. *Computational molecular evolution*. Oxford: Oxford University Press; 2006.
13. Lecointre G, Le Guyader H. *The tree of life: a phylogenetic classification*. Cambridge (MA): Belknap Press; 2006.
14. Vargas P, Zardoya R, editors. *The tree of life: evolution and classification of living organisms*. Sunderland (MA): Sinauer Associates; 2014.

Downloaded From Vet-eBooks.com

15. Farris JS. The information content of the phylogenetic system. *Syst Zool* 1979;**28**: 483−519.
16. Barta JR, Thompson RCA. What is *Cryptosporidium*? Reappraising its biology and phylogenetic affinities. *Trends Parasitol* 2006;**22**:463−8.
17. Dahlgren SS, Gjerde B. *Sarcocystis* in moose (*Alces alces*): molecular identification and phylogeny of six *Sarcocystis* species in moose, and a morphological description of three new species. *Parasitol Res* 2008;**103**:93−110.
18. Dahlgren SS, Gjerde B. The red fox (*Vulpes vulpes*) and the arctic fox (*Vulpes lagopus*) are definitive hosts of *Sarcocystis alces* and *Sarcocystis hjorti* from moose (*Alces alces*). *Parasitology* 2010;**137**:1547−57.
19. Harvey PH, Brown AJL, Maynard Smith J, Nee S, editors. *New uses for new phylogenies*. New York: Oxford University Press; 1996.
20. Sauquet H. A practical guide to molecular dating. *Comptes Rendus Palevol* 2013;**12**: 355−67.
21. Hennig W. *Phylogenetic systematics*. Urbana (IL): University of Illinois Press; 1966 [Translated by Davis DD, Zangerl R, from Hennig W. *Grundzüge einer theorie der phylogenetischen systematik*. Berlin: Deutscher Zentralverlag; 1950.].
22. Platnick NI, Cameron HD. Cladistic methods in textual, linguistic, and phylogenetic analysis. *Syst Biol* 1977;**26**:380−5.
23. Atkinson QD, Gray RD. Curious parallels and curious connections—phylogenetic thinking in biology and historical linguistics. *Syst Biol* 2005;**54**:513−26.
24. Lemey P, Derdelinckx I, Rambaut A, Van Laethem K, Dumont S, Vermeulen S, et al. Molecular footprint of drug-selective pressure in a human immunodeficiency virus transmission chain. *J Virol* 2005;**79**:11981−9.
25. Sanson GFO, Kawashita SY, Brunstein A, Briones MRS. Experimental phylogeny of neutrally evolving DNA sequences generated by a bifurcate series of nested polymerase chain reactions. *Mol Biol Evol* 2002;**19**:170−8.
26. Barta JR. Molecular approaches for inferring evolutionary relationships among protistan parasites. *Vet Parasitol* 2001;**101**:175−86.
27. Morrison DA. Evolution of the Apicomplexa: where are we now? *Trends Parasitol* 2009; **25**:375−82.
28. Morrison DA. Prospects for elucidating the phylogeny of the Apicomplexa. *Parasite* 2008; **15**:191−6.
29. Barta JR. Phylogenetic analysis of the class Sporozoea (phylum Apicomplexa Levine, 1970): evidence for the independent evolution of heteroxenous life cycles. *J Parasitol* 1989;**75**:195−206.
30. Doyle JJ. Gene trees and species trees: molecular systematics as one-character taxonomy. *Syst Bot* 1992;**17**:144−63.
31. Gatesy J, Desalle R, Whalberg N. How many genes should a systematist sample? Conflicting insights from a phylogenomic matrix characterized by replicated incongruence. *Syst Biol* 2007;**56**:355−63.
32. Corl A, Ellegren H. Sampling strategies for species trees: the effects on phylogenetic inference of the number of genes, number of individuals, and whether loci are mitochondrial, sex-linked, or autosomal. *Mol Phylogenet Evol* 2013;**67**:358−66.
33. Sapp J. *The new foundations of evolution: on the tree of life*. New York: Oxford University Press; 2009.
34. Hedges RW. The pattern of evolutionary change in bacteria. *Heredity* 1972;**28**:39−48.

Downloaded From Vet-eBooks.com

35. Ogedengbe JD, Ogedengbe ME, Hafeez MA, Barta JR. Molecular phylogenetics of eimeriid coccidia (Eimeriidae, Eimeriorina, Apicomplexa, Alveolata): a preliminary multigene and multi-genome approach. *Parasitol Res* 2015;**114**:4149–60.
36. Lemmon EM, Lemmon AR. High-throughput genomic data in systematics and phylogenetics. *Annu Rev Ecol Evol Syst* 2013;**44**(19):1–23.
37. McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT. Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol Phylogenet Evol* 2013; **66**:526–38.
38. Morrison DA. Is the tree of life the best metaphor, model or heuristic for phylogenetics? *Syst Biol* 2014;**63**:628–38.
39. Tenter AM, Barta JR, Beveridge I, Duszynski DW, Mehlhorn H, Morrison DA, et al. The conceptual basis for a new classification of the coccidia. *Int J Parasitol* 2002;**32**:595–616.
40. Morrison DA. Multiple sequence alignment for phylogenetic purposes. *Austral Syst Bot* 2006;**19**:479–539.
41. Morrison DA. A framework for phylogenetic sequence alignment. *Plant Syst Evol* 2009; **282**:127–49.
42. Morrison DA. Is multiple sequence alignment an art or a science? *Syst Bot* 2015;**40**: 14–26.
43. Morrison DA, Morgan MJ, Kelchner SA. Molecular homology and multiple sequence alignment: an analysis of concepts and practice. *Austral Syst Bot* 2015;**28**:46–62.
44. Reeck GR, de Haën C, Teller DC, Doolittle RF, Fitch WM, Dickerson RE, et al. "Homology" in proteins and nucleic acids: a terminology muddle and a way out of it. *Cell* 1987;**50**:667.
45. Morrison DA. Why would phylogeneticists ignore computerized sequence alignment? *Syst Biol* 2009;**58**:150–8.
46. Müller KF. Incorporating information from length-mutational events into phylogenetic analysis. *Mol Phylogenet Evol* 2006;**38**:667–76.
47. Ochoterena H. Homology in coding and non-coding DNA sequences: a parsimony perspective. *Plant Syst Evol* 2009;**282**:151–68.
48. Lee MSY. Unalignable sequences and molecular evolution. *Trends Ecol Evol* 2001;**16**: 681–5.
49. Gupta RS. Identification of conserved indels that are useful for classification and evolutionary studies. In: Goodfellow M, Sutcliffe I, Chun J, editors. *New approaches to prokaryotic systematics*. London: Academic Press; 2014. p. 153–82.
50. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000;**17**:540–52.
51. Kück P, Meusemann K, Dambach J, Thormann B, von Reumont BM, Wägele JW, et al. Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. *Front Zool* 2010;**7**:10.
52. Rajan V. A method of alignment masking for refining the phylogenetic signal of multiple sequence alignments. *Mol Biol Evol* 2013;**30**:689–712.
53. Bapteste E, van Iersel L, Janke A, Kelchner S, Kelk S, McInerney JO, et al. Networks: expanding evolutionary thinking. *Trends Genet* 2013;**29**:439–41.
54. Mindell DP. The tree of life: metaphor, model, and heuristic device. *Syst Biol* 2013;**62**: 479–89.
55. Xia X, Xie Z, Salemi M, Chen L, Wang Y. An index of substitution saturation and its application. *Mol Phylogenet Evol* 2003;**26**:1–7.

Downloaded From Vet-eBooks.com

56. Jermiin LS, Ho SYW, Ababneh F, Robinson J, Larkum AWD. The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. *Syst Biol* 2004;**53**:638−43.
57. Bergsten J. A review of long-branch attraction. *Cladistics* 2005;**21**:163−93.
58. Penny D, Lockhart PJ, Steel MA, Hendy MD. The role of models in reconstructing evolutionary trees. In: Scotland RW, Siebert DJ, Williams DM, editors. *Models in phylogeny reconstruction*. Oxford: Clarendon Press; 1994. p. 211−30.
59. Johnson JB, Omland KS. Model selection in ecology and evolution. *Trends Ecol Evol* 2004;**19**:101−8.
60. Gascuel O, Guindon S. Modelling the variability of evolutionary processes. In: Gascuel O, Steel M, editors. *Reconstructing evolution: new mathematical concepts and computational advances*. Oxford: Oxford University Press; 2007. p. 65−107.
61. Rodrigo A, Ewing G, Drummond A. The evolutionary analysis of measurably evolving populations using serially sampled gene sequences. In: Gascuel O, Steel M, editors. *Reconstructing evolution: new mathematical concepts and computational advances*. Oxford: Oxford University Press; 2007. p. 30−61.
62. Hasegawa N, Sugiura W, Shibata J, Matsuda M, Ren F, Tanaka H. Inferring within-patient HIV-1 evolutionary dynamics under anti-HIV therapy using serial virus samples with vSPA. *BMC Bioinformatics* 2009;**10**:360.
63. Farhat MR, Shapiro BJ, Sheppard SK, Colijn C, Murray M. A phylogeny-based sampling strategy and power calculator informs genome-wide associations study design for microbial pathogens. *Genome Med* 2014;**6**:101.
64. van Noorden R, Maher B, Nuzzo R. The top 100 papers: Nature explores the most-cited research of all time. *Nature* 2014;**514**:550−3.
65. Roger AJ, Hug LA. The origin and diversification of eukaryotes: problems with molecular phylogenetics and molecular clock estimation. *Philos Trans R Soc Lond B Biol Sci* 2006; **361**:1039−54.
66. Jeffroy O, Brinkmann H, Delsuc F, Philippe H. Phylogenomics: the beginning of incongruence? *Trends Genet* 2006;**22**:225−31.
67. Thomarat F, Vivares CP, Gouy M. Phylogenetic analysis of the complete genome sequence of *Encephalitozoon cuniculi* supports the fungal origin of microsporidia and reveals a high frequency of fast-evolving genes. *J Mol Evol* 2004;**59**:780−91.
68. Morrison DA. Pattern recognition in phylogenetics: trees and networks. In: Elloumi M, Iliopoulos CS, Wang JTL, Zomaya AY, editors. *Pattern recognition in computational molecular biology: techniques and approaches*. New York: Wiley; 2015. p. 419−38.
69. Naciri Y, Linder HP. Species delimitation and relationships: the dance of the seven veils. *Taxon* 2015;**64**:3−16.
70. Avise J, Robinson T. Hemiplasy: a new term in the lexicon of phylogenetics. *Syst Biol* 2008;**57**:503−7.
71. Nakhleh L. Computational approaches to species phylogeny inference and gene tree reconciliation. *Trends Ecol Evol* 2013;**28**:719−28.
72 Szöllösi GJ, Tannier E, Daubin V, Boussau B. The inference of gene trees with species trees. *Syst Biol* 2015;**64**:e42−62.
73. Rannala B, Yang Z. Phylogenetic inference using whole genomes. *Annu Rev Genomics Hum Genet* 2008;**9**:217−31.
74. Ren F, Tanaka H, Yang Z. A likelihood look at the supermatrix−supertree controversy. *Gene* 2009;**441**:119−25.
75. Springer MS, Gatesy J. The gene tree delusion. *Mol Phylogenet Evol* 2016;**94**:1−33.

Downloaded From Vet-eBooks.com

76. Degnan JH, Rosenberg NA. Gene tree discordance, phylogenetic inference, and the multispecies coalescent. *Trends Ecol Evol* 2009;**24**:332−40.
77. Knowles LL. Estimating species trees: methods of phylogenetic analysis when there is incongruence across genes. *Syst Biol* 2009;**58**:463−7.
78. Edwards SV, Xi Z, Janke A, Faircloth BC, McCormack JE, Glenn TC, et al. Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. *Mol Phylogenet Evol* 2016;**94**:447−62.
79. Morrison DA. Genealogies: pedigrees and phylogenies are reticulating networks not just divergent trees. *Evol Biol* 2016;**43** [in press].
80. Baum DA, Smith SD, Donovan SS. The tree-thinking challenge. *Science* 2005;**310**: 979−80.
81. Gregory TR. Understanding evolutionary trees. *Evol Educ Outreach* 2008;**1**:121−37.
82. Omland KE, Cook LG, Crisp MD. Tree thinking for all biology: the problem with reading phylogenies as ladders of progress. *BioEssays* 2008;**30**:854−67.
83. Zachos FE. Tree thinking and species delimitation: guidelines for taxonomy and phylogenetic terminology. *Mamm Biol* 2016;**81**:185−8.
84. O'Hara RJ. Population thinking and tree thinking in systematics. *Zool Scr* 1997;**26**:323−9.
85. Baum DA, Offner S. Phylogenies & tree-thinking. *Am Biol Teach* 2008;**70**:222−9.
86. Wilkinson M, McInerney JO, Hirt RP, Foster PG, Embley TM. Of clades and clans: terms for phylogenetic relationships in unrooted trees. *Trends Ecol Evol* 2007;**22**:114−5.
87. Krell F-T, Cranston PS. Which side of a tree is more basal? *Syst Entomol* 2004;**29**:279−81.
88. Crisp MD, Cook LG. Do early branching lineages signify ancestral traits? *Trends Ecol Evol* 2005;**20**:122−8.
89. Huelsenbeck JP, Bollback JP, Levine AM. Inferring the root of a phylogenetic tree. *Syst Biol* 2002;**51**:32−43.
90. Yap VB, Speed T. Rooting a phylogenetic tree with nonreversible substitution models. *BMC Evol Biol* 2005;**5**:2.
91. Sinsheimer JS, Little RJ, Lake JA. Rooting gene trees without outgroups: EP rooting. *Genome Biol Evol* 2012;**4**:709−19.
92. Williams TA, Heaps SE, Cherlin S, Nye TM, Boys RJ, Embley TM. New substitution models for rooting phylogenetic trees. *Philos Trans R Soc Lond B Biol Sci* 2015;**370**: 20140336.
93. Smith AB. Rooting molecular trees: problems and strategies. *Biol J Linn Soc* 1994;**51**: 279−92.
94. Valas RE, Bourne PE. Structural analysis of polarizing indels: an emerging consensus on the root of the tree of life. *Biol Direct* 2009;**4**:30.
95. Martin WF. Early evolution without a tree of life. *Biol Direct* 2011;**36**:6.
96. Galtier N, Daubin V. Dealing with incongruence in phylogenomic analyses. *Philos Trans R Soc Lond B Biol Sci* 2008;**363**:4023−9.
97. Bapteste E, O'Malley MA, Beiko RG, Ereshefsky M, Gogarten JP, Franklin-Hall L, et al. Prokaryotic evolution and the tree of life are two different things. *Biol Direct* 2009;**4**:34.
98. Ragan MA. Trees and networks before and after Darwin. *Biol Direct* 2009;**4**:43.
99. Sneath PHA. Cladistic representation of reticulate evolution. *Syst Zool* 1975;**24**:360−8.
100. Bremer K, Wanntorp H-E. Hierarchy and reticulation in systematics. *Syst Zool* 1979;**28**: 624−7.
101. Morrison DA. Phylogenetic networks in systematic biology (and elsewhere). In: Mohan RM, editor. *Research advances in systematic biology*. Trivandrum (India): Global Research Network; 2010. p. 1−48.

Downloaded From Vet-eBooks.com

102. Morrison DA. Phylogenetic networks: a review of methods to display evolutionary history. *Annu Res Rev Biol* 2014;**4**:1518−43.
103. Huson DH, Rupp R, Berry V, Gambette P, Paul C. Computing galled networks from real data. *Bioinformatics* 2009;**25**:i85−93.
104. Francis A, Steel M. Which phylogenetic networks are merely trees with additional arcs? *Syst Biol* 2015;**64**:768−77.
105. Morrison DA. Using data-display networks for exploratory data analysis in phylogenetic studies. *Mol Biol Evol* 2010;**27**:1044−57.
106. Minh BQ, Klaere S, von Haeseler A. Taxon selection and split diversity. *Syst Biol* 2009;**58**: 586−94.
107. Doolittle WF. The practice of classification and the theory of evolution, and what the demise of Charles Darwin's tree of life hypothesis means for both of them. *Philos Trans R Soc Lond B Biol Sci* 2009;**364**:2221−8.
108. Koonin EV. Darwinian evolution in the light of genomics. *Nucleic Acids Res* 2009;**37**: 1011−34.
109. Bray RA, Waeschenbach A, Cribb TH, Weedall GD, Dyal P, Littlewood DTJ. The phylogeny of the Lepocreadioidea (Platyhelminthes, Digenea) inferred from nuclear and mitochondrial genes: implications for their systematics and evolution. *Acta Parasitol* 2009;**54**:310−29.
110. Höglund J, Morrison DA, Mattsson JG, Engström A. Population genetics of the bovine/cattle lungworm (*Dictyocaulus viviparus*) based on mtDNA and AFLP marker techniques. *Parasitology* 2006;**133**:89−99.
111. Donoghue MJ, Baldwin BG, Li J, Winkworth RC. *Viburnum* phylogeny based on chloroplast *trn*K intron and nuclear ribosomal ITS DNA sequences. *Syst Bot* 2004;**29**:188−98.

Downloaded From Vet-eBooks.com

This page intentionally left blank

Downloaded From Vet-eBooks.com

Evolutionary Responses to Infectious Disease

9

G. Cochran, H. Harpending †
University of Utah, Salt Lake City, UT, United States

1. Introduction

Humans have a typical mammalian immune system, with three components: external barriers, the innate immune system, and the adaptive immune system. External barriers include physical barriers such as skin and mucosal surfaces as well as antibacterial secretions like lysozyme and defensins.

The second component, the innate immune system, responds quickly to attack but is not tailored to the specific attacking pathogen. It has built-in features that recognize and attack pathogens using pattern recognition receptors, which are triggered by characteristic molecular signatures associated with certain classes of disease organisms, such as the lipopolysaccharides found in the cell walls of Gram-negative bacteria. It also includes some very specific defenses against particular pathogens. The innate immune system knows that certain molecules are danger signs, but it does not learn. Its knowledge was generated by natural selection rather than individual experience, rather like the fear of snakes that is especially easily invoked in humans.

The third component, the adaptive immune system, can tailor responses to a specific pathogen and retains the ability to rapidly respond to future visitations by that pathogen. It acts as an individual immunological memory.

The human immune systems has defenses which work against practically any conceivable pathogen (the adaptive system), defenses that are pretuned against traditional classes of pathogens such as bacteria, RNA and DNA viruses, protozoa, parasitic worms, and fungi, and a number of other specialized defenses that are aimed at specific pathogens. For example, there are genes that defend against herpes simplex,[1] Epstein–Barr virus,[2] and certain dangerous strains of human papilloma virus.[3] People with two broken copies of such a gene almost inevitably have serious or lethal infections of the associated pathogen.

Some human defenses protect against regional pathogens. Mainly that means malaria, which we discuss at length later, but we also know of a built-in defense that is effective against *Trypanosoma brucei*, the cause of a common African trypanosomal infection known as nagana in livestock. The molecule is apolipoprotein L1 (APOL-1),[4] a lipoprotein: normally such molecules transport lipids, but this one also plays a role in killing trypanosomes. APOL-1 has, we suspect, played an important role in human evolution. First, it is a pretty clear signal that human evolved in sub-Saharan Africa, since that is where nagana is found. It probably says something

† Deceased.

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00009-3
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

about our preferred habitats in the early days of human evolution, since the tsetse flies carrying the disease are mainly found near rivers and lakes, in forests along watercourses, and in wooded savanna. Tsetse flies are not common in arboreal environments, and interestingly, chimpanzees do not have this defense,[5] although gorillas do.

Innate human nagana defenses are not just another signpost pointing to Africa. They also suggest possible ecological explanations for some patterns in the fossil record, for example, for the fact that Neanderthal remains have never been found in Africa. Given a long time living in areas without tsetse flies, several hundred-thousand years, Neanderthals probably had nonfunctional versions of this defense, due to relaxed selection and mutation accumulation. The loss of this defense (and most likely of other defenses against specific African pathogens) would have made it almost impossible to expand back into Africa. Since Africa was pathogen-rich compared to Europe and southwest Asia, the main Neanderthal homeland, the situation was not symmetrical. Hominids could leave Africa, but they could not go home again.

It is not just that Africa was pathogen-rich, compared to Eurasia, which might suggest that African immune systems were simply better. The optimal set of defenses depends on the environment: African defenses were better in Africa, while a different set would have been optimal for Eurasia. For one thing, Eurasia did have its own local pathogens, such as scrub typhus.[6] Second, since strong immune defenses have costs and risks (autoimmune diseases such as lupus), lower total pathogen pressure in Eurasia would have favored a turned-down, not quite as aggressive immune response. Anatomically, modern humans expanding out of Africa would certainly have developed the appropriate immune defenses, given time, but it may not have taken all that long. We now know that modern humans picked up useful variants through admixture with archaic humans such as Neanderthals[7] and Denisovans[8] that had inhabited parts of Eurasia for several hundred-thousand years, certainly enough time to adapt to local conditions. A Neanderthal variant of STAT2, an innate immune gene involved in interferon response to viral infection, is widely distributed in Eurasia and reaches a high frequency in New Guinea. Non-Africans also picked up new HLA alleles from both Neanderthals and Denisovans.[9] In retrospect, this should have been anticipated: HLA alleles are extremely varied, likely because of rare-allele advantage generated by host—pathogen coevolution. Since HLA alleles are so varied, humans are preadapted to new and different HLA variants. They are effectively interchangeable parts. This is in contrast to more typical archaic alleles—apparently there had been enough evolutionary divergence so that the typical Neanderthal variant did not quite fit in modern humans, nor ours in theirs.

Many known or suspected genetic responses to infectious disease in human are loss of function mutations, and damaged or broken genes. This is the case for nearly all the known malaria adaptations: we will discuss falciparum malaria in some detail in the following. Other protective variants that are known or suspected are also damaged versions of the wild type. Most tropical Africans have the Duffy-negative chemokine receptor, which confers protection against vivax malaria. For many years it has been thought that the deletion had no negative consequences, but Reich et al.[10] shown that the Duffy-negative allele itself causes a significant reduction in neutrophil count in people of African descent. This reduced white cell count which almost certainly has some disadvantages.

Downloaded From Vet-eBooks.com

2. Parasites as Our Friends

Of course parasites can also be helpful. Invading modern humans may have carried diseases that hit archaic humans harder during the original modern human diaspora out of Africa. This effect helped the Europeans expand into the New World. Venereal diseases would be good candidates, since they can propagate successfully at the low human densities typical of hunter-gatherers. Directly transmitted crowd diseases, on the other hand, would not have been particularly devastating to the low density archaic populations encountered in the Levant and Europe during that early expansion (as opposed to the high-density agricultural populations in the New World encountered by the Spanish).

Parasites may also have contributed to human success in hunting. We picked up three species of taenid tapeworms from African predators (another sign of our origin) several hundred-thousand years ago,[11] one from hyenas and two from lions. These tapeworms, like many other parasites, have a complex life cycle, forming cysts in herbivores (the intermediate host) and reaching maturity in the carnivorous definitive host. Obviously their interests conflict with those of their intermediate hosts, since they benefit when their host is eaten.

In such a situation, there is an evolutionary incentive for parasites to manipulate the behavior of their intermediate host, for example, by making them easier to catch.[12] This may be the case for toxoplasma, a protozoan that uses many herbivores as intermediate hosts and cats as the definitive host. Toxoplasma has been shown to cause fearlessness in rats and mice,[13] and to cause chimpanzees to develop a perverse attraction to leopards[14]: who benefits? There is evidence that *Echinococcus*, another taeniid tape worm with canid definitive hosts, increases predation on its intermediate hosts (e.g., moose). Those human tapeworms may have played an important role in human hunting success, particularly in the olden days when human weapons and hunting skills were far less sophisticated than those used by contemporary hunter-gatherers. Before agriculture, those tapeworms used wild pigs and ungulates as intermediate hosts. Now they cycle through domesticated pigs and cows, suggesting another way in which those parasites could have aided humans by fostering domestication (Ivy Smith, personal communication).

Wild boars are quite formidable, but the aurochs (the wild ancestor of domesticated cattle) was simply terrifying, being 2 m high at the shoulder and weighing over a ton. Domestication sounds difficult and dangerous, but it might have been easier if a parasite was, for its own reasons, reducing the aurochs' fear of humans.

3. Demography and Parasites

Pathogen dynamics can have a major influence on long term demographics—and the other way around of course. Pathogens typically require a minimum number of hosts in fairly close proximity (called the critical community size) in order to survive. Consider measles: it is infectious for no more than 10 days, and survivors have lifelong

Downloaded From Vet-eBooks.com

immunity. Clearly measles can only flourish in a situation where there is a steady supply of fresh, never-infected hosts, that is to say children. Because of these facts, measles has a critical community size of roughly a quarter of a million people: it could not have existed in its present form back in hunter-gatherer times, since there were no such large population concentrations.[15]

At the opposite extreme, chickenpox, after infecting children, lingers in nerve ganglia. It often recurs much later in life as shingles, which causes excruciating pain. Children can catch chickenpox from their grandfather's case of shingles. Thus, due to its persistence and ability to wait, chickenpox has a critical community size around 1000.[16]

These facts about infectious disease imply certain things about our ancestral demographics. For one thing, a population crash would have usually been followed by a boom, partly because resources become more abundant in such situations, but also because infectious diseases become less important at low population densities. A mega-crash, one in which humans had a brush with extinction, could thus have had a silver lining: one or more human-specific parasites could have gone extinct. If those parasites had imposed a heavy fitness burden, humans would have flourished after the crash. Something similar (a bottleneck in space rather than time) happens sometimes when a species colonizes a new continent—the settler population is too small to carry along key parasites and thrives to a surprising degree in its new home.

Africa is rich in human pathogens. Since we originated there, African pathogens have had a long time to adapt to humans and other primates. We mentioned that populations such as Neanderthals that spent a long time outside of Africa probably lost defenses that would have been necessary in Africa, and thus could not go back. The other side of this coin is that those vigorous defenses against African pathogens had costs, costs that were no longer necessary in cooler climates. Leaving Africa may have had substantial payoffs, first for archaic humans in Eurasia and later for anatomically modern humans.

4. Agriculture

The biggest demographic change ever experienced by humans was the population explosion made possible by the development of agriculture. Our numbers increased by factors of 50−100, which had a fundamental (and highly unpleasant) impact on human infectious disease. Pathogens that already infected humans became more common and had greater impacts on fitness, while new pathogens arose that could only spread in high-density populations—crowd diseases. We acquired most of these crowd diseases from other animals. Some originated in the animals we domesticated, while a number of others came from African primates. Some probably evolved from older human pathogens moving into newly available ecological niches.

The human genome responded to the new disease pressures, and we have observed the resulting changes in many components of the immune system. The

Downloaded From Vet-eBooks.com

35delG mutation of connexin-26 causes deafness in homozygotes, but also changes characteristics of the skin (thicker) and sweat (saltier): it may protect against infections of the skin such as erysipelas.[17] It is also a common cause of deafness in homozygotes. There is evidence of selection on a number of genes in the innate immune system such as CR-1 (a malaria defense) and some of the Toll-like receptors (TLRs) that recognize characteristic pathogen molecules. Some changes, such as the mutations causing familiar Mediterranean fever[18] and alpha-1-antitrypsin deficiency,[19] loosen protective restrictions on some of the more aggressive components of the immune system—you might compare these to unleashing the police, always a dangerous thing to do. There have been changes in the adaptive immune system as well, particularly in the major histocompatibility complex (MHC). We have recognized many of these changes because they cause serious Mendelian diseases that would hardly have reached such high frequencies unless there was some form of heterozygote advantage. Genomic scans have discovered other adaptive changes that do not have such high costs.

It is an odd fact that we seem to see fewer of these expensive disease defenses in East Asia, particularly considering only those that defend against something other than falciparum malaria. We know of no obvious reason why this should be so: conceivably it might a result of looking harder at European genetics (ascertainment bias) but right now it is something of a puzzle.

5. Some Lessons From Malaria

The case of malaria illustrates a number of general principles about the relationship between infectious disease, biological evolution, and social evolution in humans. We discuss aspects of malaria biology in some detail, but much the same story could be told for other infectious diseases of humans, for example, yellow fever. Falciparum malaria is the most serious human infectious disease and has been the strongest and best understood selective force acting on humans over the past few 1000 years. This selection pressure operated in the peoples of the Old World tropics and subtropics—but not elsewhere—and so caused those populations to diverge from the rest of humanity in some ways.[20] The most dramatic impact has been the rise to high frequency of many protective alleles. A number of those alleles (the best-studied ones) are overdominant and cause major health problems in homozygotes. The sickle-cell mutation is the most famous protective allele. Heterozygotes gain substantial protection against falciparum malaria while homozygotes suffer from a severe anemia that is usually lethal in childhood without modern medical treatment. Even so, it continues to cause substantial morbidity and mortality. It is the most common lethal mutation in humans, with a gene frequency of around 10% or more in many populations of tropical Africa.

There are a number of similar protective polymorphisms which are also disease alleles. Some change the hemoglobin molecule, either by amino acid changes (like hemoglobin C and hemoglobin E) or by changing the relative numbers of hemoglobin subunits, as in the thalassemias. Others change the red cell in different ways,

Downloaded From Vet-eBooks.com

interfering with its metabolism (glucose-6-pohophodehydrogenase (G6PD) deficiency) or altering membrane proteins (melanesian ovalocytosis). It seems likely that falciparum malaria has existed in its present form for 5000 years or less. The approximate age of some of the protective polymorphisms has been determined, and they all seem to be younger than that.[21,22] Increasingly, researchers are discovering alleles favored by malaria selection that apparently do not cause disease, not even in homozygotes. Some affect familiar targets, such as the red cell membrane (glycophorin C[23] and type O blood[24]). We also see variants of immune-system molecules such as Cd36[25] and CR-1.[26] This trend of discoveries is likely to continue, and we should eventually observe malaria-induced changes in the frequencies of many alleles, even those that have only weak effects on resistance. That is the typical pattern seen in artificial selection experiments. Strong selection for any trait other than fitness itself causes negative changes in other traits—so resistance to malaria has most likely had significant costs. Obviously we know of the costs of many that take the form of Mendelian diseases, but there are likely others as well.

Falciparum malaria's unusual virulence can be explained in part by its means of transmission. Natural selection favors low virulence in many infectious diseases that are spread directly from person to person, since immobilizing the host interferes with transmission. Since malaria is a vector-borne disease (spread by mosquitoes), a severe infection can still spread, even if the host is bedridden.[27] If high parasite blood counts increase the probability of transmission, severe infection may be a favored strategy. The other major kind of human malaria, *Plasmodium vivax*, is also mosquito-borne. It is a fairly serious disease, although much less so than falciparum. It is often found in temperate climates, where it must survive winters without active mosquitoes. In order to do so, it has the ability to hide in liver cells for long periods, in some cases for decades. Of course, this strategy would not work if the host died, which explains why vivax malaria has relatively low virulence. *Plasomodium falciparum* mainly exists in warmer climates where mosquito transmission occurs through most or all of the year, so that it can keep moving to new hosts.

Malaria has another characteristic that increases its severity. Unlike most other pathogens, malaria repeatedly switches its surface proteins. A single parasite clone has about 60 antigenic variants and thus can stay ahead of the immune system for a year or more, while greater variety in the parasite population as a whole means that a single infection does not result in lasting immunity.[28] This defensive tactic of malaria has made the development of an effective vaccine very difficult: no such vaccine is clinically available at this time.

Selection for malaria resistance in humans illustrates several key evolutionary principles: some of these are very well known, while others are not so obvious. First, it shows that adaptive evolution is a continuing process in humans, one that can cause significant changes over historical time and whose direction is not the same in every population.

This may have been especially the case over the Holocene, during which humans experienced substantial climate change, were exposed to the selective pressures associated with agriculture, and greatly increased in number.

Downloaded From Vet-eBooks.com

Malaria selection is also a clear example of convergent evolution. The protective alleles in Southeast Asia are entirely different from those in Africa: some are different mutations of genes that have produced defensive alleles in Africa (e.g., G6PD deficiency) while others involve different genes. One sees the same thing in artificial selection experiments: the phenotypic changes are similar in different lines experiencing the same selective pressures (people in both Africa and South East Asia are resistant to malaria) but the genetic details are in general different. Another point is that strong selection evokes changes in many genes, changes that are concentrated in a few metabolic paths. In this case we know of many polymorphisms that affect the red cell and hemoglobin, as well as a number that result in immunological changes. We have seen arguments that this pattern is somehow unparsimonious: one sweep might be caused by strong selective pressures, but surely not many! But in fact strong selection is likely to cause a number of sweeps—basically, every gene that significantly affects the trait under selection is a candidate for an adaptive mutation.

These convergent adaptations also show us something about the way in which advantageous alleles have spread through populations. Particular protective alleles have spread through much of sub-Saharan Africa, across New Guinea, or throughout the coastal regions of the Mediterranean—but few have managed to cross the Sahara Desert or move between India and Southeast Asia. Strong geographical barriers have prevented high-fitness alleles from spreading to all the places they would have worked—that and limited time—and thus local protective variants took their place. Evolution was faster than gene flow.

Many of these protective alleles are overdominant, since homozygotes suffer from serious disease. Overdominance means that the heterozygote has higher fitness than the homozygote: such alleles never go to fixation. A recessive lethal like sickle cell is clearly overdominant, but some of the other defensive alleles that do not cause obvious disease may also have lower fitness in homozygotes. A number of domesticated animals also have overdominant alleles that are products of recent strong selection, such as myostatin mutations in whippets and cattle. This may be a general feature of strong selection: many of the sweeping alleles generated by such selection may therefore reach maximum frequencies well under 100%.

Another interesting point comes from a simple thought experiment: there must have been a time when falciparum malaria had not existed for long and protective alleles were as yet rare. In those days, sickle cell heterozygotes (for example) should have had a larger fitness advantage, relative to the population average, than they do today, since in those days the average person had no other protective alleles. Today, on the other hand, someone in Africa who does not carry the sickle cell allele is likely to have a number of other protective alleles—alpha thalassemia, G6PD deficiency, and so on. Africans who do not carry sickle cell are still far more resistant to falciparum malaria than northern Europeans or Amerindians.[29] Hence, the fitness advantage of being a sickle cell carrier (which was as high as 20% in recent centuries) must have been even larger thousands of years ago. This means that the rate of growth, and the equilibrium frequency, if overdominant, of every allele that protected against

Downloaded From Vet-eBooks.com

malaria slowed down as time passed, as the population acquired more and more resistance to malaria from other alleles. This effect can also stop a selective sweep short of fixation.

We think that falciparum malaria has had another interesting effect on human evolution, in that it often kept populations below the Malthusian limit—that is, kept population density below the level at which resource limitations would have stopped further growth. In a Malthusian situation, resources are short and individuals compete for them. Selection in that situation favors efficient use of available resources, which would involve improvements in metabolic and work efficiency—basically, farmers who can plow more acres per calorie. It also favors paternal investment. At the limit, you end up with hard-working peasant couples (both father and mother) who can just barely manage to raise enough food for themselves and the two children who replace them in the next generation.

In a sub-Malthusian ecology, where factors like disease and/or violence keep the population well below the subsistence limit, selection pushes in a different direction. Here the limiting factor might be health rather than wealth. Disease resistance in a mate could be more valuable than land, hence a father's genetic quality might be important than his provisioning ability. In much of Africa today, women do most of the farm work: this low level of paternal investment is only possible when resources are plentiful. Female self-sufficiency combined with a high value placed on genetic quality favors polygyny (multiple wives), since man's genes are more important than his wages. Polygyny is more common in West Africa than anywhere else.

6. Disease and Standard of Living in Preindustrial Societies: A Simple Model

We can elaborate the role of disease in shaping human cultural diversity with a simple model. Disease in a population that would otherwise be Malthusian, that is to say resource limited, can have the effect of reducing the population size, leading to an increase in the standard of living of those who remain. A familiar example is the prosperity and high wages in Europe following the massive human die off following the great plague epidemics.

We start with a small group of 1000 colonizers in an empty environment. Initially the population is at such a low density that there is no competition for resources among people. Births and deaths occur at a constant per-person rate. There is no age structure—no youth, nor old people—so everyone is subject to the same rates; these assumptions make algebraic models easy and they reflect well what happens in more realistic (but more complicated) models. Plausible generic values for low-technology human populations are 50 births per 1000 people per year and 30 deaths per 1000 people per year. The difference, 20 per 1000 per year, is the intrinsic growth rate, 20 per 1000 or 2 per 100, 2% per year. In the absence of any limitation the population grows according to this rate exactly like money at

Downloaded From Vet-eBooks.com

compound interest. After a generation of 25 years the population size would be $1000 \times (1.02)^{25}$ or 1640 people. This population would double in about 35 years and would double slightly more than 14 times in 500 years to an implausibly large size of nearly 20 million people.

The customary models in population geneticists focus on gene frequency change, and mean fitness, population growth rate, is normalized away in the equations for genetic change. Here we need to acknowledge the demographic consequences of gene change and retain the mean-fitness parameter in order to study the interaction of demography and genetics. The mean fitness of 1.02 per year (or 1.64 per generation) occurs at low population densities, but gradually declines as the population grows to the carrying capacity. At this limit, the mean fitness is just 1.0, and the population remains stationary.

What are the long term implications of this modest rate of growth? The rate of 2% per year is commonplace among human populations yet a growth of 2,000,000% over 500 years seems and is outlandish. Early in the process resources would become scarce and the rate of growth would slow. Assuming the initial colony occupied 100 square miles, the expanded population after five centuries would need to occupy nearly 2,000,000 (2 million) square miles, about the area of Argentina or Kazakhstan. This is explosive growth in historical time but it corresponds to everyday population growth today in many low-technology societies. We know that over the long period from the modern human diaspora out of Africa about 45,000 years ago to the industrial revolution about 200 years ago human numbers grew but at long term rates far below our modest 2% per year. On this long time scale they hardly grew at all. It is likely that most of the time populations were growing at rates like our 2%, perhaps slower, but that there were frequent catastrophic events like wars, famines, and plagues that cut population sizes back.

7. Population Limitation

There is a convenient and standard way to make a model of population limited by resources, called the logistic model. This may not be very accurate but it is simple and, given our poor understanding of detailed dynamics, more than good enough. The idea is that there is some carrying capacity K of the environment. Populations below the carrying capacity in size can grow while populations above the carrying capacity decline until they reach K. If we write P_t for population in some generation t and P_{t+1} for population the following generation then simple population growth like compound interest, called geometric, follows this formula:

$$P_{t+1} = P_t \times (1 + R)$$

where R is just the intrinsic growth rate and $(1 + R)$ is the mean fitness. We write $R = 0.641$ in the expression for the intrinsic growth rate since a rate of 2% per year

Downloaded From Vet-eBooks.com

corresponds to growth of 64.1% per generation. The logistic model specifies that the growth rate R is damped by the current ratio of population to carrying capacity:

$$P_{t+1} = P_t \times \left(1 + R \times \left(1 - \frac{P_t}{K}\right)\right)$$

In an empty environment without intraspecific competition population P is much less than carrying capacity K and population growth is almost the same as the simple geometric case. But as population increases the ratio P/K becomes significant, growth slows down, and eventually stops when population reaches carrying capacity, that is, when $P = K$. If the carrying capacity of the environment into which our population moved were 10,000 people then the population would grow at a decreasing rate to reach 10,000.

What if the carrying capacity is not static but increases with the number of people? For example, we might imagine that more people bring more farmland under cultivation so that K itself changes. It turns out[30] that nothing much changes if the increase in carrying capacity K is proportionally less than the increase in population P as would happen if the best land were cleared first while lower and lower quality land were subsequently brought under cultivation. The population still approaches a (new larger) carrying capacity so that as equilibrium is approached population P is equal to carrying capacity. The end result is that the standard of living, by which we mean the ratio of resources to people K/P, is still unity. There are more people but they are not living any better than they did before the new land was cleared. Such a population, limited by resources, is referred to as a Malthusian population.

An interesting variant of this model is to introduce a new source of mortality, perhaps disease or warfare.[31] In areas of central Africa with high levels of falciparum malaria, malaria's fitness cost may be around 25%: that means that with malaria an average individual will leave 25% fewer living descendants one generation later. With a growth rate of 20 per 1000 per year an average individual has 1.64 daughters one generation later. If malaria now decreases fitness by 25% the average individual will only have 75% of 1.64 or 1.23 daughters one generation later. In terms of annual rates the malaria cuts population growth from 2% to 0.8% per year. (Notice that we count only daughters since our model is of a simple population that does not take into account sexual reproduction.)

Now we can consider the fixed carrying capacity K and examine the consequences for the population and for individual well-being. The algebraic model now becomes (writing M for the extra density-independent death rate, from malaria in our example but also likely to be from violence and local warfare):

$$P_{t+1} = P_t \times \left(1 + R - \frac{RP_t}{K} - M\right)$$

Downloaded From Vet-eBooks.com

We can find the equilibrium population, that is, the population that would remain unchanging in this environment with the extra mortality. We simply set P_{t+1} equal to P_t, rearrange some terms, and find that the new equilibrium is at

$$P = K \times \frac{(R - M)}{R}.$$

We substitute our assumed values, an intrinsic growth rate R of 0.64 and an extra mortality rate M of 0.25 to obtain

$$\frac{P}{K} = \frac{0.39}{0.64} \sim 0.61.$$

The population now equilibrates at 61% of the old carrying capacity. A more interesting way to summarize what we have found by manipulating the logistic model is in terms of the standard of living, where a value of 1 means the bare subsistence minimum compatible with life and the maintenance and population size and a value of, say, 5, means that there is five times the subsistence minimum amount of resources available to the average person. In our model population the standard of living is the reciprocal of 0.61 or 1.6. There is more than half again as many resources per person as there were before malaria appeared. What this means on the ground is that people do not have to work very hard to get enough to eat, that there is fruit on the trees for plucking, and that there are not great labor demands on anyone. Those who survive the malaria enjoy a much higher standard of living. Fig. 9.1 is a

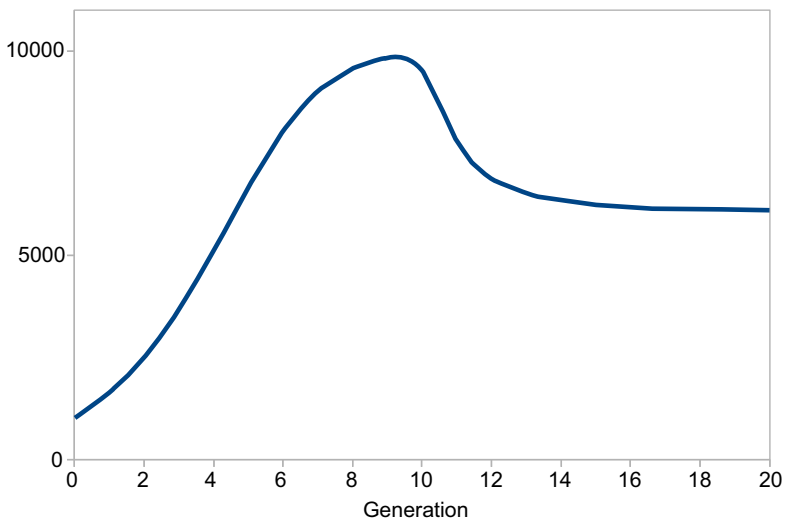


Figure 9.1 Model population size over time of a population of 1000 introduced into an empty area with a carrying capacity of 10,000. After 250 years (10 generations) falciparum malaria appears, and population size quickly drops to about 6500. Generation is on the *x*-axis, population size on the *y*-axis.

Downloaded From Vet-eBooks.com

simulation of this process using plausible numbers for a low-technology human population. The population grows 10-fold from 1000 to 10,000 people in 10 generations, then quickly shrinks to the new equilibrium size of 6100 people after the introduction of falciparum malaria.

Gregory Clark[32] points out that the medieval English had a higher standard than the medieval Japanese because there was much more sewage and filth in England, and so a heavier burden of disease. This extra disease translated, as in our malaria example, to a lower population density and higher standard of living.

What are the social consequences of this new disease for low-technology human populations? The most important immediate consequence is that there are plentiful resources for everyone and so, following the nature of the creature, males withdraw from subsistence work as they find that they can simply parasitize women for food. In much of central Africa, Oceania, and the Americas the result is or has been societies in which men do not do anything very useful and women provision themselves, their children, and the men. The euphemism in economics for this kind of society is "female farming system." Left free of the demands of subsistence the men start hanging out together, perhaps even all moving into a village men's house (not so uncommon in Africa). This leads to local and regional raiding and warfare, and an entrenched culture of local violence.

8. Disease, Mating, and Reproductive Strategy

Several decades ago Hamilton and Zuk[33] showed a correlation in North American passerines between parasite burden and gaudiness. Their model was that a slowly changing parasite load leads to parent−offspring correlations in ducks in parasite resistance leading to mating preferences for bright colors as signals of that resistance. Subsequent literature suggests that a similar phenomenon occurs in human societies.[34,35]

The underlying logic is clear enough, much of it similar to that of the earlier discussion of endemic malaria. Human females, like all mammals, may obtain provisioning from a male for herself and her offspring. This is the pattern in many settled agricultural societies where male subsistence labor is necessary for successful reproduction: these are so-called dad societies. On the other hand in societies that are far below the Malthusian limit, female mate preferences are more likely to favor males other than good providers. For example, in these societies, often characterized by chronic local raiding and warfare, males protect females from other males. In places with high endemic disease loads then, as in ducks, disease resistance can be heritable so that females prefer to mate with males with "good genes" rather than with males who are "good providers." Of course if females are selecting males with good genes rather than males who are good providers then this is an open door for polygyny. A peasant farmer would have great difficulty provisioning several families but no difficulty at all simply mating with several females. Traditional societies of tropical Africa are indeed mostly polygynous.

Downloaded From Vet-eBooks.com

There seem to be several important ecological routes to quasi-stable high standard of living non-Malthusian societies. One is warfare: in many well documented cases deaths by violence approach a quarter to one half of all deaths with the result that human densities remain below any purely subsistence population limit.[36] Much of highland New Guinea appears to be a classic example of this cultural ecosystem. Nothing much seems to have changed there in many millennia. The rough broken terrain contributes to the persistence of this system since the terrain makes it nearly impossible for an effective constabulary to suppress the chronic violence.

Another route to such a society is through a high burden of endemic disease as in the malaria example. Malaria is of course the classic case but yellow fever, hookworm, and many others should have similar effects. Much of sub-Saharan Africa is described by economists as "female farming systems," a euphemism for societies where men essentially parasitize women for subsistence while they commit more effort than males do elsewhere to subtle and not-so-subtle male−male competition. As females prefer males who appear healthy (for their "good genes") they may be selection in males to accommodate this preference, that is, there should be sexual selection for appearance. Several recently described myostatin mutants in Africa[37] are probably recently evolved signals of male quality.

The worldwide fall of fertility rates following the industrial revolution in northern nations suggests that a stable non-Malthusian world is attainable without the unpleasantness and misery of violence and infectious disease. Meanwhile it is important to understand that a disease like falciparum malaria not only causes much human misery directly, it also leaves in its wake damaging genetic traces that may take hundreds of generations to dissipate. It also leaves in its wake a social order likely not so well suited for modern industrial society. While tough fierce physically attractive males may be favored in a social system where there are adequate resources for females to do all the provisioning, these same are not going to do so well in a subsistence ecology that demands hard agricultural labor and actively sanctions violent behavior.

9. Prosperity and the Postindustrial Era Mortality Decline

It has become apparent in the last decade that evolution in humans is an ongoing process that is even speeding up in the face of drastic cultural changes and the large number of humans on earth, each of whom is a potential target for mutations, including favorable mutations.[38] Clark[32] has proposed that genetic change during the millennium or so before the industrial revolution led to essentially a new version of humans that made the revolution possible. The greatest change in human economic history since the origins of agriculture, the industrial revolution of around 1800, released our species from Malthusian constraints as income growth suddenly outstripped population growth.

Downloaded From Vet-eBooks.com

This revolution in human society was accompanied by many changes, and no one has a very clear idea of how they are related to each other. Clark, whose focus is on Great Britain, emphasizes these changes as follows:

- A decline in propensity to violence, especially male violence. In the nations of Europe homicide rates fell by one to two orders of magnitude in the millennium before 1800.[39,40] In many preindustrial societies violent males enjoyed a reproductive advantage through greater access to mates, but that advantage turned into a severe disadvantage in settled agricultural societies with effective constabularies.
- Declining interest rates reflecting declining time preference. People were more and more inclined to delay gratification.
- An increasing affinity for work.
- A strong correlation between wealth and reproductive success of males.

However, at the same time, according to Clark, there were other equally profound changes:

- A striking mortality decline, the cause of which is not well understood. Civil engineering and vaccination are often suggested as causes of this decline but the evidence is not very clear. The decline may also reflect in part genetic adaptation to new kinds of infectious disease.
- Birth rates fell drastically. The fall of birth rates lagged the fall in death rates by several decades. This is the so-called "demographic transition" from high mortality and high fertility to low mortality and low fertility. This led to the relationship between wealth and fertility to reverse, as it is today in industrial societies. Today wealthier people have fewer surviving offspring, and this reversal was the immediate precursor of the popular eugenics movement of the late 19th and early 20th centuries.

What is the role, if any, of the mortality decline in this seismic shift in the nature of society? A possibility is that the decline in pressure from infectious disease freed up much of the genome to evolve in new directions determined by the new social environment. Most of the genetic adaptations to infectious disease that we (think that we) understand involve major or minor damage to genes and to individuals. We discussed the sickle cell adaptation to malaria earlier with its high, purely genetic death toll on homozygotes, but there are many parallel adaptations to falciparum malaria that are known and almost certainly many more that we do not yet understand. In aggregate they must impose a large genetic burden on populations with a history of living with malaria.

As the prevalence of infectious disease declined in pre- and postindustrial societies there may have been widespread relaxation of the selection maintaining these damaging genetic polymorphisms, with the effect of releasing these constraints on the genome and facilitating selection to move phenotypes in different, more favorable directions in the phenotype space.

References

1. Sancho-Shimizu V, Zhang S-Y, Abel L, Tardieu M, Rozenberg F, Jouanguy E, et al. Genetic susceptibility to herpes simplex virus 1 encephalitis in mice and humans. *Curr Opin Allergy Clin Immunol* December 2007;**7**(6):495−505.
2. Rigaud S, Fondanèche MC, Lambert N, Pasquier B, Mateo V, Soulas P, et al. XIAP deficiency in humans causes an X-linked lymphoproliferative syndrome. *Nature* 2006;**444**(7115):110−4.

Downloaded From Vet-eBooks.com

3. Ramoz N, Rueda L-A, Bouadjar B, Montoya L-S, Orth G, Favre M. Mutations in two adjacent novel genes are associated with epidermodysplasia verruciformis. *Nat Genet* December 2002;**32**(4):579−81.
4. Pays E, Vanhollebeke B, Vanhamme L, Paturiaux-Hanocq F, Nolan DP, Pérez-Morga D. The trypanolytic factor of human serum. *Nat Rev Microbiol* 2006;**4**(6):477−86.
5. Puente XS, Gutiérrez-Fernández A, Ordóñez GR, Hillier LW, López-Otín C. Comparative genomic analysis of human and chimpanzee proteases. *Genomics* December 2005;**86**(6): 638−47.
6. Traub R, Wisseman CL. Review article: the ecology of chigger-borne rickettsiosis (scrub typhus). *J Med Entomol* 1974;**11**(3):237−303.
7. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft sequence of the Neandertal genome. *Science* May 7, 2010;**328**(5979):710−22.
8. Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* December 23, 2010; **468**(7327):1053−60.
9. Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E. Evidence for archaic adaptive introgression in humans. *Nat Rev Genet* 2015;**16**(6):359−71.
10. Reich D, Nalls MA, Kao WHL, Akylbekova EL, Tandon A, Patterson N, et al. Reduced neutrophil count in people of African descent is due to a regulatory variant in the Duffy antigen receptor for chemokines gene. *PLoS Genet* January 30, 2009;**5**(1):e1000360.
11. Hoberg EP. Phylogeny of Taenia: species definitions and origins of human parasites. *Parasitol Int* 2006;**55**(Suppl. 1):S23−30.
12. Lagrue C, Poulin R. Manipulative parasites in the world of veterinary science: implications for epidemiology and pathology. *Vet J* April 2010;**184**(1):9−13.
13. Berdoy M, Webster JP, Macdonald DW. Fatal attraction in rats infected with *Toxoplasma gondii*. *Proc Biol Sci* August 7, 2000;**267**(1452):1591−4.
14. Poirotte C, Kappeler PM, Ngoubangoye B, Bourgeois S, Moussodji M, Charpentier MJ. Morbid attraction to leopard urine in Toxoplasma-infected chimpanzees. *Curr Biol* 2016; **26**(3):R98−9.
15. Black FL. Measles endemicity in insular populations: critical community size and its evolutionary implication. *J Theor Biol* July 1966;**11**(2):207−11.
16. Black FL, Hierholzer WJ, Pinheiro FDK, Evans AS, Woodall JP, Opton EM, et al. Evidence for persistence of infectious agents in isolated human populations. *Am J Epidemiol* 1974; **100**(3):230.
17. Meyer CG, Amedofu GK, Brandner JM, Pohland D, Timmann C, Horstmann RD. Selection for deafness? *Nat Med* 2002;**8**:1332−3.
18. Aksentijevich I, Torosyan Y, Samuels J, Centola M, Pras E, Chae JJ, et al. Mutation and haplotype studies of familial Mediterranean fever reveal new ancestral relationships and evidence for a high carrier frequency with reduced penetrance in the Ashkenazi Jewish population. *Am J Hum Genet* 1999;**64**(4):949−62.
19. Lomas DA. The selective advantage of α1-antitrypsin deficiency. *Am J Respir Crit Care Med* 2006;**173**(10):1072−7.
20. Pennington R, Gatenbee C, Kennedy B, Harpending H, Cochran G. Group differences in proneness to inflammation. *Infect Genet Evol* 2009;**9**(6):1371−80.
21. Ohashi J, Naka I, Patarapotikul J, Hananantachai H, Brittenham G, Looareesuwan S, et al. Extended linkage disequilibrium surrounding the hemoglobin E variant due to malarial selection. *Am J Hum Genet* June 2004;**74**(6):1198−208.
22. Saunders MA, Slatkin M, Garner C, Hammer MF, Nachman MW. The extent of linkage disequilibrium caused by selection on G6PD in humans. *Genetics* 2005;**171**(3):1219.

Downloaded From Vet-eBooks.com

23. Maier AG, Duraisingh MT, Reeder JC, Patel SS, Kazura JW, Zimmerman PA, et al. Plasmodium falciparum erythrocyte invasion through glycophorin C and selection for Gerbich negativity in human populations. *Nat Med* January 2003;**9**(1):87−92.
24. Rowe JA, Opi DH, Williams TN. Blood groups and malaria: fresh insights into pathogenesis and identification of targets for intervention. *Curr Opin Hematol* November 2009; **16**(6):480−7.
25. Pain A, Urban BC, Kai O, Casals-Pascual C, Shafi J, Marsh K, et al. A non-sense mutation in Cd36 gene is associated with protection from severe malaria. *Lancet* 2001;**357**(9267): 1502−3.
26. Cockburn IA, Mackinnon MJ, O'Donnell A, Allen SJ, Moulds JM, Baisor M, et al. A human complement receptor 1 polymorphism that reduces *Plasmodium falciparum* rosetting confers protection against severe malaria. *Proc Natl Acad Sci USA* January 6, 2004;**101**(1):272−7.
27. Ewald PW. *Evolution of infectious disease*. USA: Oxford University Press; 1994.
28. Scherf A, Lopez-Rubio JJ, Riviere L. Antigenic variation in *Plasmodium falciparum*. *Annu Rev Microbiol* October 2008;**62**(1):445−70.
29. Mackinnon MJ, Mwangi TW, Snow RW, Marsh K, Williams TN. Heritability of malaria in Africa. *PLoS Med* 2005;**2**(12):e340.
30. Cohen JE. How many people can the earth support? *Sciences* 1995;**35**(6):18−23.
31. Armstrong RA, Gilpin ME. Evolution in a time-varying environment. *Science* 1977; **195**(4278):591−2.
32. Clark G. A farewell to alms : a brief economic history of the world [Internet]. In: *The Princeton economic history of the western world*. Princeton: Princeton University Press; 2007. Available from: http://www.loc.gov/catdir/toc/ecip0715/2007015166.html. http://www.loc.gov/catdir/enhancements/fy0814/2007015166-b.html. http://www.loc.gov/catdir/enhancements/fy0814/2007015166-d.html.
33. Hamilton WD, Zuk M. Heritable true fitness and bright birds: a role for parasites? *Science* 1982;**218**(4570):384−7.
34. Gangestad SW, Buss DM. *Pathogen prevalence and human mate preferences*. Elsevier; 1993.
35. Low BS. Marriage systems and pathogen stress in human societies. *Am Zool* 1990;**30**(2): 325−40.
36. Keeley LH. *War before civilization*. 1996.
37. Saunders MA, Good JM, Lawrence EC, Ferrell RE, Li WH, Nachman MW. Human adaptive evolution at myostatin (GDF8), a regulator of muscle growth. *Am J Hum Genet* 2006;**79**:1089.
38. Hawks J, Wang ET, Cochran GM, Harpending HC, Moyzis RK. Recent acceleration of human adaptive evolution. *Proc Natl Acad Sci USA* 2007;**104**(52):20753.
39. Eisner M. Modernization, self-control and lethal violence. The long-term dynamics of European homicide rates in theoretical perspective. *Br J Criminol* 2001;**41**(4):618−38.
40. Eisner M. Long-term historical trends in violent crime. *Crime Justice* 2003:83−142.

Downloaded From Vet-eBooks.com

Infectious Disease Genomics

10

Y.-T. Liu
University of California San Diego, La Jolla, CA, United States

1. Introduction

The history and development of infectious disease genomics are closely associated with the Human Genome Project (HGP).[1] A series of important discussions about the HGP were made from 1985 to 1986,[1,2] which led to the appointment of a special National Research Council (NRC) committee by the National Academy of Sciences to address the needs and concerns, such as its impact, leadership, and funding sources. The committee recommended that the United States begin the HGP in 1988.[3] They emphasized the need for technological improvements in the efficiency of gene mapping, sequencing, and data analysis capabilities. In order to understand potential functions of human genes through comparative sequence analyses, they also advised that the HGP must not be restricted to the human genome and should include model organisms including mouse, bacteria, yeast, fruit fly, and worm. In the meantime, the Office of Technology Assessment (OTA) of the US Congress also issued a similar report to support the HGP.[4] In 1990, the Department of Energy (DOE) and the National Institutes of Health (NIH) jointly presented an initial 5-year plan for the HGP.[5] In October 1993, the Sanger Center/Institute (Hinxton, UK) was officially open to join the HGP. The cost of DNA sequencing was about $2 to $5 per base in 1990 and the initial aim was to reduce the costs to less than $0.50 per base before large-scale sequencing.[5] The sequencing cost gradually declined during the subsequent years. In 2004, the National Human Genome Research Institute (NHGRI) challenged scientists to achieve a $100,000 human genome (3 Gb/haploid genome) by 2009 and a $1000 genome by 2014 to meet the need of genomic medicine. In early 2014, Illumina announced that the company would begin producing a new system to deliver full coverage human genomes for less than $1,000.[6]

The first complete genome to be sequenced was the phiX174 bacteriophage (5.4 kb) by Sanger's group in 1977.[7] The complete genome sequence of SV40 polyomavirus (5.2 kb) was published in 1978.[8,9] The human Epstein–Barr virus (170 kb) genome was determined in 1984.[10] The first completed free-living organism genome was *Haemophilus influenza* (1.8 Mb), sequenced through a whole-genome shotgun approach in 1995.[11] The second sequenced bacterial genome, *Mycoplasma genitalium* (600 kb), was completed in less than 1 month in the same year using the same approach.[12] The DOE was the first to start a microbial genome program (MGP) as a companion to its HGP in 1994.[13] The initial focus was on nonpathogenic microbes. Along with the development of the HGP, there was exponential growth of the number of completely sequenced free-living organism genomes. The Fungal Genome Initiative

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00010-X
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

(FGI)[14] was established in 2000 to accelerate the slow pace of fungal genome sequencing since the report of the genome of *Saccharomyces cerevisiae* in 1996.[15] One of the major interests was to sequence organisms that are important in human-health and commercial activities. With the explosion in the number of sequenced genomes, thanks to the development of next generation−sequencing methods, many genome-based studies have become popular. Compared to 6 years ago when only 1100 completed genome projects were documented, the GOLD (Genomes OnLine Database) contains information for 67,879 genome-sequencing projects, of which 7210 were completed, as of August 2015.[16,17]

The genomes of human malaria parasite *Plasmodium falciparum* and its major mosquito vector *Anopheles gambiae* were published in 2002.[18,19] Historically, the effort to sequence the malaria genome began in 1996 by taking advantage of a clone derived from laboratory-adapted strain.[20] Notably, many parasites have complex life cycles that involve both vertebrate and invertebrate hosts and are difficult to maintain in the laboratory. Few other important human pathogenic parasites, such as trypanosomes,[21,22] *Leishmania*,[23] and schistosomes, [24,25] have been either completely or partially sequenced.[26,27] In the meantime, the genome sequence of *Aedes aegypti*, the primary vector for yellow fever and dengue fever, was published in 2007.[28] The genome size (1376 Mb) of this mosquito vector is about 5 times larger than the previously sequenced genome of the malaria vector *A. gambiae*. About 50% of the genome consists of transposable elements. In 2010, the genome sequence of the body louse (*Pediculus humanus humanus*), an obligatory parasite of humans and the main vector of epidemic typhus (*Rickettsia prowazekii*), relapsing fever (*Borrelia recurrentis*), and trench fever (*Bartonella quintana*), was reported.[29] Its 108 Mb genome is the smallest among the known insect genomes. Subsequently, more vector genomes have been published.[30−32] Genome-sequencing projects for other important human disease vectors are in progress.[33,34] These include *Culex pipiens* (mosquito vector of West Nile virus), and *Ixodes scapularis* (tick vector of Lyme disease, *Babesia* and *Anaplasma*). The challenge to sequence the genome of an insect vector is much greater than a microbe. For example, the genome of ticks was estimated to be between 1 and 7 Gb and may have a significant proportion of repetitive DNA sequences, which may be a problem for genome assembly.[35] Furthermore, the evolutionary distances among insect species may also affect homology-based gene predictions.

It is as important to understand the sequence diversity within a species as to perform a de novo sequencing of a reference genome from the perspective of human health. This is true for both hosts and pathogens.[36,37] The goal of the 1000 Genomes Project is to find most genetic variants that have frequencies of at least 1% in the human populations studied.[38] One of the similar efforts for human pathogens is the NIH Influenza Genome Sequencing Project. When this project began in November 2004, only seven human influenza H3N2 isolates had been completely sequenced and deposited in the GenBank database.[39,40] As of May 2010, more than 5000 human and avian isolates had been completely sequenced, including the 1918 “Spanish” influenza virus.[41] Databases for human immunodeficiency virus (HIV) and hepatitis C virus have also been established.

Downloaded From Vet-eBooks.com

While most human studies of microbes have focused on the disease-causing organisms, interest in resident microorganisms has also been growing. In fact, it has been estimated that the human body is colonized by at least 10 times more prokaryotic and eukaryotic microorganisms than the number of human cells.[42] It was suggested to have "the 2nd human genome project" to sequence the human microbiome.[43] Highly variable intestinal microbial flora among normal individuals has been well documented.[44–46] Therefore, the Human Microbiome Project (HMP) was initiated by the NIH in late 2007.[47] The analysis and data of 242 healthy adults at 15 (for males) or 18 (for females) body sites over 22 months were published in 2012.[48]

The completed or ongoing genome projects (Table 10.1) provide enormous opportunities for the discovery of novel vaccines and drug targets against human pathogens as well as the improvement of diagnosis and discovery of infectious agents and the development of new strategies for invertebrate vector control. Specific examples are

Table 10.1 Completed or Ongoing Genome Projects

General
NCBI[106] (http://www.ncbi.nlm.nih.gov/sites/genome)
ENSEMBL[107] (http://www.ensemblgenomes.org/)
JCVI[108] (http://cmr.jcvi.org/)
GOLD[16] (http://www.genomesonline.org)
Sanger Pathogen genomics (http://www.sanger.ac.uk/Projects/Pathogens/)
GeMInA (genomic metadata for infectious Agents)[109,110] (http://gemina.igs.umaryland.edu)
Bacteria
Human Microbiome Project[111] (http://www.hmpdacc.org/)
Fungi
Fungal Genome Initiative (FGI) (http://www.broadinstitute.org/science/projects/fungal-genome-initiative)
Parasites
Eukaryotic pathogens[27] (http://EuPathDB.org)
Parasite genome projects (http://www.pasteur.fr/recherche/unites/tcruzi/minoprio/genomics/parasites.htm)
Invertebrate vectors
VectorBase[33,34] (http://www.vectorbase.org)
Viruses
Influenza virus[112] (http://www.ncbi.nlm.nih.gov/genomes/FLU/)
HIV (http://www.hiv.lanl.gov/)
HCV (http://hcv.lanl.gov/)

Downloaded From Vet-eBooks.com

provided to illustrate how the information provided by various genome projects may help achieve the goal of promoting human health.

2. Vaccine Target

Meningococcal isolates produce one of 13 antigenically distinct capsular polysaccharides, but only five (A, B, C, W135, and Y) are commonly associated with disease.[49] The polysaccharide capsule is important for meningococci to escape from complement-mediated killing. While conventional vaccines consisting of the conjugation of capsular polysaccharides to carrier proteins for meningococcus serogroups A, C, Y, and W-135 have been clinically successful, the same approach failed to produce clinically useful vaccine for serogroup B (MenB). The capsule polysaccharide ($\alpha 2-8$-*N*-acetylneuraminic acid) of MenB is identical to human polysialic acid, therefore is poorly immunogenic.[50] Alternatively, vaccines consisting of outer-membrane vesicles (OMVs) have been successfully developed to control MenB outbreaks in areas where epidemics are dominated by one particular strain.[51–54] The most significant limitation of this type of vaccine is that the immune response is strain specific, mostly directed against the porin protein, PorA, which varies substantially in both expression level and sequence across strains.[55,56]

With the completion of the genome sequence of a virulent MenB strain, a "reverse vaccinology" approach was applied for the development of a universal MenB vaccine by Novartis.[55,57,58] Through bioinformatic searching for surface exposed antigens, which may be the most suitable vaccine candidates due to their potential to be readily recognized by the immune system, 570 open reading frames (ORFs) were selected from a total of 2158 ORFs of the MC58 genome. Eventually, five antigens were chosen as the vaccine components based on a series of criteria including the ability of candidates to be expressed in *Escherichia coli* as recombinant proteins (350 candidates), the confirmation of surface exposure by immunological analyses, the ability of induced protective antibodies in experimental animals (28 candidates), and the conservation of antigens within a panel of diverse meningococcal strains, primarily the disease-associated MenB strains.[55,58,59] The vaccine formulation consists of an fHBP-GNA2091 fusion protein, a GNA2132-GNA1030 fusion protein, NadA, and OMVs from the New Zealand MeNZB vaccine strain, which contains the immunogenic PorA. Initial phase II clinical results in adults and infants showed that this vaccine could induce a protective immune response against three diverse MenB strains in 89–96% of subjects following three vaccinations and 93–100% after four vaccinations.[59] This vaccine (Bexsero) has been approved in the USA and in more than 30 other countries.[60]

3. New Drug Discovery

Natural products, especially microbial secondary metabolites, are important source of bioactive compounds. Actinomycetes have been a main source of natural-product

Downloaded From Vet-eBooks.com

discovery in bacteria. Consequently, the high rediscovery rate of known compounds and scaffolds were inevitable with activity-based screening. Genome mining of gene clusters that produce secondary metabolites have been a new approach to overcome this problem. For example, an antibiotic, clostrubin, was discovered through searching novel compounds from *Clostridium beijerinckii* due to the presence of several cryptic gene clusters for secondary metabolite biosynthesis.[61]

Genome mining starts with a genome-wide search for highly conserved members of the required biosynthesis gene cluster. Computational programs that support the prediction of operons help to assign boundaries of newly identified biosynthesis gene clusters. A large-scale, high-throughput genome mining for the genetic potential for producing phosphonic acids by screening more than 10,000 actinomycetes has been achieved in 2015.[62] It was believed that phosphonates would have greater potential to become pharmaceuticals, with a past commercialization rate of 15% (3/20), such as fosfomycin, compared to the 0.1% average for natural products as a whole.[63,64] In addition, bioinformatical discovery of phosphonate biosynthetic loci has been well established, as all but two previously characterized phosphonate biosynthetic pathways start with phosphoenolpyruvate (PEP) mutase that is encoded by *pepM*. Among 10,000 actinomycetes, only 278 strains were confirmed to have *pepM* by polymerase chain reaction (PCR) screening and genome sequencing. A diverse collection of phosphonate biosynthetic gene clusters were identified within these strains. Remarkably, 55 out of the 64 distinct clusters would direct the synthesis of unknown compounds. Characterization of strains within five of these groups resulted in discovery of argolaphos, and other interesting compounds, including valinophos, and phosphonocystoximate. Argolaphos showed broad-spectrum antibacterial activity against *Salmonella typhimurium*, *E. coli*, and *Staphylococcus aureus*.

4. Drug Target

Targeting an essential pathway is a necessary but not sufficient requirement for an effective antimicrobial agent.[65] Identification of essential genes in a completely sequenced genome has been actively pursued with various approaches.[66,67] The indispensable fatty acid synthase (FAS) pathway in bacteria has been regarded as a promising target for the development of antimicrobial agents.[68] The subcellular organization of the fatty acid biosynthesis components is different between mammals (type I FAS) and bacteria (dissociated type II FAS), which raises the likelihood of host specificity of the targeting drugs. Comparison of the available genome sequences of various species of prokaryotes reveals highly conserved FAS II systems suggesting that the antimicrobial agent can be broad spectrum.[69] In addition, through computational analyses, new members of the FAS II system have been discovered in different bacterial species.[70,71] One of the protein components in this system, FabI, is the target

Downloaded From Vet-eBooks.com

of an antituberculosis drug isonizid and a general antibacterial and antifungal agent, triclosan.[72–74]

Through a systematic screening of 250,000 natural product extracts, a Merck team identified a potent and broad-spectrum antibiotic, platensimycin, which is derived from *Streptomyces platensis* and a selective FabF/B inhibitor in FAS II system.[75] Treatment with platensimycin eradicated *S. aureus* infection in mice. Platensimycin did not have cross-resistance to other antibiotic-resistant strains in vitro, including methicillin-resistant *S. aureus*, vancomycin-intermediate *S. aureus*, and vancomycin-resistant enterococci. No toxicity was observed using a cultured human cell line and the activity of platensimycin was not affected by the presence of human serum in this study. However, the FAS II system appears to be dispensable for another Gram-positive bacterium, *Streptococcus agalactiae*, when exogenous fatty acids are available, such as in human serum.[65,76] The susceptibility to inhibitors targeting the FAS II system indicates heterogeneity in fatty acid synthesis or in acquiring exogenous fatty acids among Gram-positive pathogens.[76] Comparative genomic approaches may be useful to identify and develop a strategy to target the salvage pathway for *S. agalactiae*. Alternatively, similar approaches as described earlier for MenB vaccine may also be applied for *S. agalactiae* (Group B *Streptococcus*).[77]

5. Therapeutic Response and Drug Resistance

Emergence of drug-resistant malaria to chloroquine in 1950s and sulfadoxine–pyrimethamine in 1960s occurred from western Cambodia to the Greater Mekong subregion (GMSR, including Cambodia, Lao, Myanmar, Thailand, and Vietnam) and to Africa. The finding of artemisinin-resistant malaria in Cambodia and GMSR raised a concern regarding the global spread of these parasites. While a number of studies, including population genetics and laboratory-based investigations were conducted, no reliable molecular marker was identified until the major breakthrough reported in early 2014.[78] Clinical artemisinin resistance has been defined as a reduction of parasite-clearance rate, which is expressed as an increase of parasite-clearance half-life, or a persistence of microscopically detectable parasites 3 days after artemisinin-based combination therapy (ACT). Although artemisinin was thought to have broad-stage specificity against malaria throughout the life cycle, it was showed that artemisinin-resistant parasites only had decrease of artemisinin susceptibility at ring stages, which was demonstrated by the ring-stage survival assay ($RSA_{0–3\ h}$).[79]

An in vitro laboratory-based approach was conducted at a time when population-based genome-wide association studies (GWAS) did not clearly identify the genes responsible for artemisinin resistance.[78] For 5 years, an artemisinin-resistant F32-ART5 parasite line was selected by culturing an artemisinin-sensitive F32-Tanzania clone under a dose-escalating, 125-cycle regimen of artemisinin. Eight mutations in seven genes were eventually selected from the result based on

Downloaded From Vet-eBooks.com

whole-genome sequence analysis F32-ART5 and F32-TEM (its sibling clone cultured without artemisinin) at 460× and 500× average nucleotide coverage, respectively. To examine whether these in vitro selected mutations were associated with artemisinin resistance in Cambodia, sequence polymorphism in all seven genes were analyzed from 49 culture-adapted clinical isolates related to their $RSA_{0-3\ h}$. Only polymorphisms of a gene, K13-propeller, showed a significant association with $RSA_{0-3\ h}$ survival rates. In total, four mutant alleles, each harboring a single nonsynonymous SNP (Y493H, R539T, I543T, and C580Y) within a kelch repeat of the C-terminal K13-propeller domain were identified. To confirm that K13-propeller polymorphism is a molecular marker of clinical artemisinin resistance, parasite-clearance half-lives in patients were correlated with their K13 alleles. Of the 150 patients, 72 carried parasites with a wild-type allele and the others carried parasites with only one of the three single nonsynonymous SNPs in the K13-propeller: C580Y ($n = 51$), R539T ($n = 6$), and Y493H ($n = 21$). The parasite-clearance half-life in patients with wild-type parasites is significantly shorter (median 3.30 h) than those with these three mutant alleles (median 6.28−7.19 h). Subsequently, clinical studies have validated the association between K13 propeller mutations and artemisinin resistance.[80−82]

6. Vector Control

Early mathematical model for malaria control suggested that the most vulnerable element in the malaria cycle was survivorship of adult female mosquitos.[83,84] Therefore, insect control is an important part of reducing transmission. The use of DDT as an indoor residual spray in the global malaria eradication program from 1957 to 1969 has reduced the population at risk of malaria to about 50% by 1975 compared with 77% in 1900.[83,85] Engineering genetically modified mosquitoes refractory to malaria infection appeared to be an alternative approach,[86] given the environmental impact of DDT and the emergence of insecticide-resistant insects. The Vector Biology Network (VBN) was formed in 1989 and had proposed a 20-year plan with the WHO in 2001 to achieve three major goals: (1) to develop basic tools for the stable transformation of anopheline mosquitoes by the year 2000, (2) to engineer a mosquito incapable of carrying the malaria parasite by 2005, and (3) to run controlled experiments to test how to drive the engineered genotype into wild mosquito populations by 2010.[87−89] While some proof-of-concept experiments have been achieved for the first two aims in 2002 when the *A. gambiae* genome was completely sequenced,[90,91] the progress has been relatively slow.[92]

Genomic loci of the *A. gambiae* responsible for *P. falciparum* resistance have been identified through surveying a mosquito population in a West African malaria transmission zone.[93] A candidate gene, *Anopheles Plasmodium*-responsive leucine-rich repeat 1 (APL1) was discovered. Subsequently, other resistant genes have also been identified.[94,95] Studying the genetic basis of resistance to malaria parasites and immunity of the mosquito vector will be important to control malaria transmission.[96]

Downloaded From Vet-eBooks.com

7. Clinical Application

Perhaps the most immediate impact of a completely sequenced pathogen genome is for infectious disease diagnosis. The information may be of great importance to the public health when a newly emerged or reemerged pathogen is discovered. A few examples will be described.

A novel swine-origin influenza A virus (S-OIV) emerged in the spring of 2009 in Mexico and subsequently was discovered in specimens from two unrelated children in the San Diego area in mid-April 2009.[97,98] Those samples were positive for influenza A but negative for both human H1 and H3 subtypes. The complete genome sequence and a real-time PCR−based diagnostic assay were released to the public in late April. The outbreak evolved rapidly and WHO declared the highest Phase 6 worldwide pandemic alert on June 11, 2009. S-OIV has three genome segments (HA, NP, and NS) from the classic North American swine (H1N1) lineage, two segments (PB2 and PA) from the North American avian lineage, one segment (PB1) from the seasonal H3N2, and most notably, two segments (NA and M) from the Eurasian swine (H1N1) lineage.[98] With the available influenza genome database, diagnostic assays to distinguish previous seasonal H1N1, H3N2, and S-OIV can be easily accomplished.[99]

A comprehensive pathogen genome database is not only useful for infectious disease diagnosis but also for novel pathogen discovery.[100] Homologous sequences within the same family or among different family members are important for new pathogen identification even with the advent of third generation−sequencing technology.[101] *De novo* pathogen discovery may also be complicated by coexisting microorganisms, such as commensal bacteria in the human body. Without prior knowledge of these microorganisms, one may be misled.

In 2003, a microarray-based assay, designated Virochip, was used to help discover the SARS conoronavirus.[102] The Virochip contained the most highly conserved 70mer sequences from every fully sequenced reference viral genome in GenBank. The computational search for conservation was performed across all known viral families. A microarray hybridized with a reaction derived from a viral isolate cultivated from a SARS (severe acute respiratory syndrome) patient revealed that the strongest hybridizing array elements belong to families Astroviridae and Coronaviridae. Alignment of the oligonucleotide probes having the highest signals showed that all four hybridizing oligonucleotides from the Astroviridae and one oligonucleotide from avian infectious bronchitis virus, an avian coronavirus, shared a core consensus motif spanning 33 nucleotides. Interestingly, it had been known previously through bioinformatics analyses that this sequence is present in the 3′ UTR of all astroviruses, avian infectious bronchitis virus, and an equine rhinovirus.[103] Therefore, a new member of the coronavirus was identified through the unique hybridizing pattern and subsequent confirmations.

The finding of the seventh human oncogenic virus, Merkel cell polyomavirus (MCV)[104] in 2008 is another example of why conserved sequences are important for novel pathogen discovery. MCV is the etiological agent of Merkel cell carcinoma

Downloaded From Vet-eBooks.com

(MCC), which is a rare but aggressive skin cancer of neuroendocrine origin. Two cDNA libraries derived from MCC tumors were subjected to high-throughput sequencing by a next-generation Roche/454 sequencer. Nearly 400,000 sequence reads were generated. The majority (99.4%) of the sequences derived from human origin were removed from further analyses. Only one of the remaining 2395 cDNA was homologous to the T antigen of two known polyomaviruses. One additional cDNA was subsequently identified to be part of the MCV sequence when the complete viral sequence was known. Later analyses showed that 80% (8/10) of the MCC had integrated MCV in the human genome. Monoclonal viral integration was revealed by the patterns of Southern blot analysis. Only 8−16% of control tissues had low copy number of MCV infection.

In 2015, an interesting and unexpected discovery of the malignant transformation of *Hymenolepis nana*, a human tape worm, in a human host has been reported by conventional and next generation−sequencing approaches.[104a] Initially, examination of a 41-year-old HIV-infected man revealed extensive lymphadenopathy. *H. nana* eggs and *Blastocystis hominis* cysts were found in stool. The disease progressed to death despite antiparasitic and antiretroviral treatment. Histological examination of biopsied lymph nodes revealed proliferative cells with overt malignant features. They were monomorphic with morphologic features characteristic of stem cells (a high nucleus-to-cytoplasm ratio). However, the small cell size (<10) suggested infection with an unfamiliar, possibly unicellular, eukaryotic organism. Infection with a plasmodial slime mold rather than *H. nana* was considered because of the prominent syncytia formation and the primitive appearance of the atypical cells but lack of architecture identifiable as tapeworm tissue. PCR screening suggested that these cells were *H. nana*. Next generation−genome sequencing and comparative analysis revealed *H. nana* variants harboring mutations typically found in cancer.

As of 2016, next generation−sequencing technologies are gradually being applied for diagnosis and monitoring of infectious diseases, including genotypic resistance testing, direct detection of unknown disease-associated pathogens without culture, investigation of microbial population diversity in the host, and strain typing.[105] However, promising, next generation−sequencing approaches for clinical diagnosis require further improvements for automation, standardization of technical and bioinformatic procedures, and other practical issues, such as costs and turnaround time.

8. Conclusion

While we can expect that the efforts of a variety of genome projects may improve human health, the socioeconomic issues that are not discussed in this chapter may be substantial. In addition, the tremendous amount of information derived from these projects will also pose a challenge for scientists as well nonscientists to follow and understand.

Downloaded From Vet-eBooks.com

References

1. Watson JD. The human genome project: past, present, and future. *Science* 1990;**248**: 44−9.
2. Dulbecco R. A turning point in cancer research: sequencing the human genome. *Science* 1986;**231**:1055−6.
3. NRC. *Mapping and sequencing the human genome*. 1988. http://www.nap.edu/catalog.php?record_id=1097.
4. OTA. *Mapping our genes—genome projects: how big? How fast?*. 1988. http://www.ornl.gov/sci/techresources/Human_Genome/publicat/OTAreport.pdf.
5. DHHS and DOE. *Understanding our genetic inheritance, the U.S. Human genome project: the first five years: fiscal years 1991−1995*. 1990. http://www.ornl.govf/sci/techresources/Human_Genome/project/5yrplan/summary.shtml.
6. Hayden EC. The $1,000 genome. *Nature* 2014;**507**:294−5.
7. Sanger F, Air GM, Barrell BG, et al. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 1977;**265**:687−95.
8. Reddy VB, Thimmappaya B, Dhar R, et al. The genome of simian virus 40. *Science* 1978; **200**:494−502.
9. Fiers W, Contreras R, Haegemann G, et al. Complete nucleotide sequence of SV40 DNA. *Nature* 1978;**273**:113−20.
10. Baer R, Bankier AT, Biggin MD, et al. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. *Nature* 1984;**310**:207−11.
11. Fleischmann RD, Adams MD, White O, et al. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* 1995;**269**:496−512.
12. Smith HO. Hisotry of microbial genomics. In: Fraser CM, Read TD, Nelson KE, editors. *Microbial genomes*. Totowa, NJ: Humana; 2004. p. 3−16.
13. DOE. *Microbial genome program*. 2009. http://microbialgenomics.energy.gov/mgp.shtml.
14. Campbell IG, Russell SE, Choong DY, et al. Mutation of the PIK3CA gene in ovarian and breast cancer. *Cancer Res* 2004;**64**:7678−81.
15. Goffeau A, Barrell BG, Bussey H, et al. Life with 6000 genes. *Science* 1996;**274**(546): 63−7.
16. Liolios K, Chen IM, Mavromatis K, et al. The Genomes on Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2010;**38**:D346−54.
17. Reddy TB, Thomas AD, Stamatis D, et al. The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. *Nucleic Acids Res* 2015;**43**:D1099−106.
18. Gardner MJ, Hall N, Fung E, et al. Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* 2002;**419**:498−511.
19. Holt RA, Subramanian GM, Halpern A, et al. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 2002;**298**:129−49.
20. Hoffman SL, Bancroft WH, Gottlieb M, et al. Funding for malaria genome sequencing. *Nature* 1997;**387**:647.
21. El-Sayed NM, Myler PJ, Bartholomeu DC, et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 2005;**309**:409−15.
22. Berriman M, Ghedin E, Hertz-Fowler C, et al. The genome of the African trypanosome *Trypanosoma brucei*. *Science* 2005;**309**:416−22.

Downloaded From Vet-eBooks.com

23. Ivens AC, Peacock CS, Worthey EA, et al. The genome of the kinetoplastid parasite, *Leishmania major*. *Science* 2005;**309**:436−42.
24. Berriman M, Haas BJ, LoVerde PT, et al. The genome of the blood fluke *Schistosoma mansoni*. *Nature* 2009;**460**:352−8.
25. Consortium. The *Schistosoma japonicum* genome reveals features of host-parasite interplay. *Nature* 2009;**460**:345−51.
26. Brindley PJ, Mitreva M, Ghedin E, Lustigman S. Helminth genomics: the implications for human health. *PLoS Negl Trop Dis* 2009;**3**:e538.
27. Aurrecoechea C, Brestelli J, Brunk BP, et al. EuPathDB: a portal to eukaryotic pathogen databases. *Nucleic Acids Res* 2010;**38**:D415−9.
28. Nene V, Wortman JR, Lawson D, et al. Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science* 2007;**316**:1718−23.
29. Kirkness EF, Haas BJ, Sun W, et al. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc Natl Acad Sci USA* 2010;**107**(27):12168−73.
30. Neafsey DE, Waterhouse RM, Abai MR, et al. Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 *Anopheles mosquitoes*. *Science* 2015;**347**:1258522.
31. Mesquita RD, Vionette-Amaral RJ, Lowenberger C, et al. Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. *Proc Natl Acad Sci USA* 2015;**112**:14936−41.
32. International Glossina Genome I. Genome sequence of the tsetse fly (*Glossina morsitans*): vector of African trypanosomiasis. *Science* 2014;**344**:380−6.
33. Lawson D, Arensburger P, Atkinson P, et al. VectorBase: a data resource for invertebrate vector genomics. *Nucleic Acids Res* 2009;**37**:D583−7.
34. Megy K, Hammond M, Lawson D, Bruggner RV, Birney E, Collins FH. Genomic resources for invertebrate vectors of human pathogens, and the role of VectorBase. *Infect Genet Evol* 2009;**9**:308−13.
35. Pagel Van Zee J, Geraci NS, Guerrero FD, et al. Tick genomics: the Ixodes genome project and beyond. *Int J Parasitol* 2007;**37**:1297−305.
36. Feero WG, Guttmacher AE, Collins FS. The genome gets personal−almost. *JAMA* 2008; **299**:1351−2.
37. Alcais A, Abel L, Casanova JL. Human genetics of infectious diseases: between proof of principle and paradigm. *J Clin Invest* 2009;**119**:2506−14.
38. Kaiser J. DNA sequencing. A plan to capture human diversity in 1000 genomes. *Science* 2008;**319**:395.
39. Fauci AS. Race against time. *Nature* 2005;**435**:423−4.
40. Ghedin E, Sengamalay NA, Shumway M, et al. Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature* 2005;**437**:1162−6.
41. Taubenberger JK, Reid AH, Lourens RM, Wang R, Jin G, Fanning TG. Characterization of the 1918 influenza virus polymerase genes. *Nature* 2005;**437**:889−93.
42. Savage DC. Microbial ecology of the gastrointestinal tract. *Annu Rev Microbiol* 1977;**31**: 107−33.
43. Relman DA, Falkow S. The meaning and impact of the human genome sequence for microbiology. *Trends Microbiol* 2001;**9**:206−8.
44. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial community variation in human body habitats across space and time. *Science* 2009;**326**:1694−7.
45. Eckburg PB, Bik EM, Bernstein CN, et al. Diversity of the human intestinal microbial flora. *Science* 2005;**308**:1635−8.

Downloaded From Vet-eBooks.com

46. Turnbaugh PJ, Hamady M, Yatsunenko T, et al. A core gut microbiome in obese and lean twins. *Nature* 2009;**457**:480−4.
47. Relman DA. Microbiology learning about who we are. *Nature* 2012;**486**:194−5.
48. Human Microbiome Project C. Structure, function and diversity of the healthy human microbiome. *Nature* 2012;**486**:207−14.
49. Lo H, Tang CM, Exley RM. Mechanisms of avoidance of host immunity by Neisseria meningitidis and its effect on vaccine development. *Lancet Infect Dis* 2009;**9**:418−27.
50. Finne J, Bitter-Suermann D, Goridis C, Finne U. An IgG monoclonal antibody to group B meningococci cross-reacts with developmentally regulated polysialic acid units of glycoproteins in neural and extraneural tissues. *J Immunol* 1987;**138**:4402−7.
51. Bjune G, Hoiby EA, Gronnesby JK, et al. Effect of outer membrane vesicle vaccine against group B meningococcal disease in Norway. *Lancet* 1991;**338**:1093−6.
52. Sierra GV, Campa HC, Varcacel NM, et al. Vaccine against group B Neisseria meningitidis: protection trial and mass vaccination results in Cuba. *NIPH Ann* 1991;**14**: 195−207. discussion 8−10.
53. Jackson C, Lennon DR, Sotutu VT, et al. Phase II meningococcal B vesicle vaccine trial in New Zealand infants. *Arch Dis Child* 2009;**94**:745−51.
54. Boslego J, Garcia J, Cruz C, et al. Efficacy, safety, and immunogenicity of a meningococcal group B (15:P1.3) outer membrane protein vaccine in Iquique, Chile. Chilean National Committee for Meningococcal Disease. *Vaccine* 1995;**13**:821−9.
55. Pizza M, Scarlato V, Masignani V, et al. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* 2000;**287**:1816−20.
56. Martin SL, Borrow R, van der Ley P, Dawson M, Fox AJ, Cartwright KA. Effect of sequence variation in meningococcal PorA outer membrane protein on the effectiveness of a hexavalent PorA outer membrane vesicle vaccine. *Vaccine* 2000;**18**:2476−81.
57. Tettelin H, Saunders NJ, Heidelberg J, et al. Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. *Science* 2000;**287**:1809−15.
58. Giuliani MM, Adu-Bobie J, Comanducci M, et al. A universal vaccine for serogroup B meningococcus. *Proc Natl Acad Sci USA* 2006;**103**:10834−9.
59. Rinaudo CD, Telford JL, Rappuoli R, Seib KL. Vaccinology in the genome era. *J Clin Invest* 2009;**119**:2515−25.
60. Bruno L, Cortese M, Rappuoli R, Merola M. Lessons from reverse vaccinology for viral vaccine design. *Curr Opin Virol* 2015;**11**:89−97.
61. Pidot S, Ishida K, Cyrulies M, Hertweck C. Discovery of clostrubin, an exceptional polyphenolic polyketide antibiotic from a strictly anaerobic bacterium. *Angew Chem Int Ed Engl* 2014;**53**:7856−9.
62. Ju KS, Gao J, Doroghazi JR, et al. Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes. *Proc Natl Acad Sci USA* 2015;**112**: 12175−80.
63. Berdy J. Thoughts and facts about antibiotics: where we are now and where we are heading. *J Antibiot* 2012;**65**:385−95.
64. Metcalf WW, van der Donk WA. Biosynthesis of phosphonic and phosphinic acid natural products. *Annu Rev Biochem* 2009;**78**:65−94.
65. Brinster S, Lamberet G, Staels B, Trieu-Cuot P, Gruss A, Poyart C. Type II fatty acid synthesis is not a suitable antibiotic target for Gram-positive pathogens. *Nature* 2009;**458**: 83−6.
66. Ji Y, Zhang B, Van SF, et al. Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. *Science* 2001;**293**:2266−9.

Downloaded From Vet-eBooks.com

67. Hutchison CA, Peterson SN, Gill SR, et al. Global transposon mutagenesis and a minimal *Mycoplasma genome*. *Science* 1999;**286**:2165−9.
68. Wright HT, Reynolds KA. Antibacterial targets in fatty acid biosynthesis. *Curr Opin Microbiol* 2007;**10**:447−53.
69. Zhang YM, Marrakchi H, White SW, Rock CO. The application of computational methods to explore the diversity and structure of bacterial fatty acid synthase. *J Lipid Res* 2003;**44**:1−10.
70. Heath RJ, Rock CO. A triclosan-resistant bacterial enzyme. *Nature* 2000;**406**:145−6.
71. Marrakchi H, Choi KH, Rock CO. A new mechanism for anaerobic unsaturated fatty acid formation in *Streptococcus pneumoniae*. *J Biol Chem* 2002;**277**:44809−16.
72. Levy CW, Roujeinikova A, Sedelnikova S, et al. Molecular basis of triclosan activity. *Nature* 1999;**398**:383−4.
73. Zhang YM, White SW, Rock CO. Inhibiting bacterial fatty acid synthesis. *J Biol Chem* 2006;**281**:17541−4.
74. Banerjee A, Dubnau E, Quemard A, et al. inhA, a gene encoding a target for isoniazid and ethionamide in *Mycobacterium tuberculosis*. *Science* 1994;**263**:227−30.
75. Wang J, Soisson SM, Young K, et al. Platensimycin is a selective FabF inhibitor with potent antibiotic properties. *Nature* 2006;**441**:358−61.
76. Balemans W, Lounis N, Gilissen R, et al. Essentiality of FASII pathway for *Staphylococcus aureus*. *Nature* 2010;**463**:E3. discussion E4.
77. Maione D, Margarit I, Rinaudo CD, et al. Identification of a universal Group B streptococcus vaccine by multiple genome screen. *Science* 2005;**309**:148−50.
78. Ariey F, Witkowski B, Amaratunga C, et al. A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria. *Nature* 2014;**505**:50−5.
79. Witkowski B, Khim N, Chim P, et al. Reduced artemisinin susceptibility of *Plasmodium falciparum* ring stages in western Cambodia. *Antimicrob Agents Chemother* 2013;**57**:914−23.
80. Ashley EA, Dhorda M, Fairhurst RM, et al. Spread of artemisinin resistance in *Plasmodium falciparum* malaria. *N Engl J Med* 2014;**371**:411−23.
81. Miotto O, Amato R, Ashley EA, et al. Genetic architecture of artemisinin-resistant *Plasmodium falciparum*. *Nat Genet* 2015;**47**:226−34.
82. Tun KM, Imwong M, Lwin KM, et al. Spread of artemisinin-resistant *Plasmodium falciparum* in Myanmar: a cross-sectional survey of the K13 molecular marker. *Lancet Infect Dis* 2015;**15**:415−21.
83. Enayati A, Hemingway J. Malaria management: past, present, and future. *Annu Rev Entomol* 2010;**55**:569−91.
84. Macdonald G. *The epidemiology and control of malaria*. Oxford: Oxford Univ. Press; 1957.
85. Hay SI, Guerra CA, Tatem AJ, Noor AM, Snow RW. The global distribution and population at risk of malaria: past, present, and future. *Lancet Infect Dis* 2004;**4**:327−36.
86. Curtis CF. Possible use of translocations to fix desirable genes in insect pest populations. *Nature* 1968;**218**:368−9.
87. Beaty BJ, Prager DJ, James AA, et al. From Tucson to genomics and transgenics: the vector biology network and the emergence of modern vector biology. *PLoS Negl Trop Dis* 2009;**3**:e343.
88. Morel CM, Toure YT, Dobrokhotov B, Oduola AM. The mosquito genome−a breakthrough for public health. *Science* 2002;**298**:79.
89. Alphey L, Beard CB, Billingsley P, et al. Malaria control with genetically manipulated insect vectors. *Science* 2002;**298**:119−21.

Downloaded From Vet-eBooks.com

90. Catteruccia F, Nolan T, Loukeris TG, et al. Stable germline transformation of the malaria mosquito *Anopheles stephensi*. *Nature* 2000;**405**:959–62.
91. Ito J, Ghosh A, Moreira LA, Wimmer EA, Jacobs-Lorena M. Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. *Nature* 2002;**417**:452–5.
92. Marshall JM, Taylor CE. Malaria control with transgenic mosquitoes. *PLoS Med* 2009;**6**: e20.
93. Riehle MM, Markianos K, Niare O, et al. Natural malaria infection in *Anopheles gambiae* is regulated by a single genomic control region. *Science* 2006;**312**:577–9.
94. Povelones M, Waterhouse RM, Kafatos FC, Christophides GK. Leucine-rich repeat protein complex activates mosquito complement in defense against *Plasmodium* parasites. *Science* 2009;**324**:258–61.
95. Blandin SA, Wang-Sattler R, Lamacchia M, et al. Dissecting the genetic basis of resistance to malaria parasites in *Anopheles gambiae*. *Science* 2009;**326**:147–50.
96. Severo MS, Levashina EA. Mosquito defenses against *Plasmodium* parasites. *Curr Opin Insect Sci* 2014;**3**:30–6.
97. CDC. Swine influenza A (H1N1) infection in two children–Southern California, March–April 2009. *MMWR Morb Mortal Wkly Rep* 2009;**58**:400–2.
98. Dawood FS, Jain S, Finelli L, et al. Emergence of a novel swine-origin influenza A (H1N1) virus in humans. *N Engl J Med* 2009;**360**:2605–15.
99. Lu Q, Zhang XQ, Pond SL, Reed S, Schooley RT, Liu YT. Detection in 2009 of the swine origin influenza A (H1N1) virus by a subtyping microarray. *J Clin Microbiol* 2009;**47**: 3060–1.
100. Liu YT. A technological update of molecular diagnostics for infectious diseases. *Infect Disord Drug Targets* 2008;**8**:183–8.
101. Munroe DJ, Harris TJ. Third-generation sequencing fireworks at Marco Island. *Nat Biotechnol* 2010;**28**:426–8.
102. Wang D, Urisman A, Liu YT, et al. Viral discovery and sequence recovery using DNA microarrays. *PLoS Biol* 2003;**1**:E2.
103. Jonassen CM, Jonassen TO, Grinde B. A common RNA motif in the 3' end of the genomes of astroviruses, avian infectious bronchitis virus and an equine rhinovirus. *J Gen Virol* 1998;**79**(Pt 4):715–8.
104. Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science* 2008. http://dx.doi.org/10.1126/science.1152586.
104a. Muehlenbachs A, Bhatnagar J, Agudelo CA, et al. Malignant transformation of hymenolepis nana in a human host. *N Engl J Med* 2015;**373**(19):1845–52. http://dx.doi.org/10.1056/NEJMoa1505892.
105. Lefterova MI, Suarez CJ, Banaei N, Pinsky BA. Next-generation sequencing for infectious disease diagnosis and management: a report of the association for molecular pathology. *J Mol Diagn* 2015;**17**:623–34.
106. Sayers EW, Barrett T, Benson DA, et al. Database resources of the National Center for Biotechnology information. *Nucleic Acids Res* 2010;**38**:D5–16.
107. Kersey PJ, Lawson D, Birney E, et al. Ensembl Genomes: extending Ensembl across the taxonomic space. *Nucleic Acids Res* 2010;**38**:D563–9.
108. Davidsen T, Beck E, Ganapathy A, et al. The comprehensive microbial resource. *Nucleic Acids Res* 2010;**38**:D340–5.
109. Ecker DJ, Sampath R, Willett P, et al. The Microbial Rosetta Stone Database: a compilation of global and emerging infectious microorganisms and bioterrorist threat agents. *BMC Microbiol* 2005;**5**:19.

Downloaded From Vet-eBooks.com

110. Schriml LM, Arze C, Nadendla S, et al. GeMInA, genomic metadata for infectious agents, a geospatial surveillance pathogen database. *Nucleic Acids Res* 2010;**38**:D754−64.
111. Nelson KE, Weinstock GM, Highlander SK, et al. A catalog of reference genomes from the human microbiome. *Science* 2010;**328**:994−9.
112. Bao Y, Bolotov P, Dernovoy D, et al. The influenza virus resource at the National Center for Biotechnology information. *J Virol* 2008;**82**:596−601.

Downloaded From Vet-eBooks.com

This page intentionally left blank

Downloaded From Vet-eBooks.com

Proteomics and Host–Pathogen Interactions: A Bright Future?

11

D.G. Biron [1], D. Nedelkov [2], D. Missé [3], P. Holzmuller [4]
[1]Laboratoire Microorganismes: Génome et Environnement, UMR CNRS/UBP/UDA 6023, Aubière Cedex, France; [2]Arizona State University, Tempe, AZ, United States; [3]Laboratoire MIVEGEC, UMR CNRS 5290/IRD 224/UM, Montpellier Cedex, France; [4]UMR CIRAD-INRA Contrôle des maladies exotiques émergentes (CMAEE), Montpellier Cedex, France

1. Introduction

Living organisms are constantly exposed to pathogens. In any environment, a molecular war begins when a host encounters a pathogen. In many host–pathogen associations, the molecular war is in progress a long time ago. Nevertheless, a disease as an outcome of a pathogen attack remains an exception rather than a rule. Most host species have acquired strategies by selective pressure to mislead the pathogen and to win the fight during their cross talk (i.e., molecular dialogue). However, many pathogen species have acquired strategies by selective pressure to bypass the host defenses to win the molecular war and to ensure the completion of their life cycle. Pathogens remain a significant threat to any host species. Critical to the mitigation of this threat is the ability to rapidly detect, respond to, treat, and contain the pathogen transmission. Since many centuries, some scientific fields (i.e., agroecology, evolutionary ecology, evolutionary medicine, biochemistry, microbiology, medicine, veterinary medicine, immunology, and molecular biology) have surveyed host–parasite interactions to improve our understanding of pathogenic diseases and to prevent pathogen transmission in host populations.

During the course of human history, pathogenic diseases have seriously affected many societies worldwide. In Europe, one of the most dramatic disease events was the great plague pandemic of the mid-14th century.[1,2] Notably, pathogenic diseases are a leading cause of premature death in the world. Pathogenic diseases result from an intimate relationship between a host and a pathogen which involves molecular "cross talk." Clearly, elucidation of this complex molecular dialogue between host and pathogen is desirable in order to improve our understanding of pathogen virulence, to develop pathogen-specific host biomarkers, and to define novel therapeutic and vaccine targets. Proteomics applications to decipher host–parasite interactions are in their infancy in spite of important technological and scientific advances since the post-genomic era, and should lead to new insights on host specificity and on the evolution of pathogen virulence. In this chapter, we present the interest of proteomics to survey host–pathogen interactions, a synthetic review of previous proteomics studies, the pitfalls of the current approach in surveys, new conceptual approaches to decipher

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00011-1
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

host–parasite interactions, a new avenue to decipher the cross-talk diversity involved in trophic interactions in a habitat (i.e., the population proteomics), and a 5-year view for future prospects on proteomics and host–pathogen interactions.

2. Interest of Proteomics to Study Host–Pathogen Interactions

Since the start of the genomic era in the early 1990s, many parasitologists and molecular biologists are confident that complete sequencing of the genome of the partners in host–pathogen associations for pathogens with simple life cycle (i.e., one host) and in host–vector–pathogen associations for pathogens with complex life cycle (i.e., at least two hosts) will enable the total understanding of the molecular mechanisms involved in most of the pathogenic diseases and will contribute to find new drugs for treating them[3,4]; insufficient progress has been achieved in the control of such diseases as malaria and sleeping sickness, despite decades of intensive genomic projects on host–pathogen interactions, vaccines, and chemotherapeutics. Pathogens continue to be a major cause of morbidity and mortality in humans and domestic livestock, especially in developing countries.[5–9]

Until now, many parasitologists and molecular biologists have focused their studies on DNA analyses based on the central dogma of molecular biology—that is to say, the general pathway for the expression of genetic information stored in DNA. Although the basic blueprint of life is encoded in DNA, the execution of the genetic plan is carried out by the activities of proteins. The fabric of biological diversity is therefore protein based, and natural selection acts at the protein level.[10] At the end of the 20th century, it had become clear to many parasitologists and molecular biologists that knowing genome sequences, while technically mandatory, was not in itself enough to fully understand complex biological events, such as the immune response of a host to a pathogen infection or the molecular strategies used by pathogens to thwart the host defenses during their interaction.[11–15]

The evolution of any given species has tremendously increased complexity at the level of pre- (gene splicing, mRNA editing) and posttranslational (phosphorylation, glycosylation, acetylation, and so on) gene–protein interaction. The genomics era has revealed that: (1) DNA sequences may be "fundamental," but can provide little information on the dynamic processes within and between host and a parasite during their physical and molecular interaction[11,12]; (2) the correlation between the expressed "transcriptome" (i.e., total mRNA transcription pattern) and the levels of translated proteins is poor[16–18]; and (3) a single gene can produce different protein products.[13,14,18] Moreover, the structure, function, abundance, and even the number of proteins in an organism cannot yet be predicted from the DNA sequence alone.[11,17,19] Also, posttranslational modifications, such as phosphorylation and glycosylation, are often extremely important for the function of many proteins, although most of these modifications cannot yet be predicted from genomic or mRNA sequences.[17] Thus, the biological phenotype of an organism is not directly related to its genotype (i.e., DNA sequences).

Downloaded From Vet-eBooks.com

Epigenetic systems control and modify gene expression. Almost all the elements of epigenetic control systems are proteins.[19] The cells of an organism are reactive systems in which information flows not only from genes to proteins but in the reverse direction as well.[3] The proteome is the genome operating system by which the cells of an organism react to environmental signals.[19] It comprises an afferent arm, the cytosensorium (i.e., many cellular proteins are sensors, receptors, and information transfer units from environmental signals) and an efferent arm, the cytoeffectorium (i.e., in cells, reaction of the genome via regulation of either individual proteins or a group of proteins in response to environmental changes).

Proteomics is the study of the proteome. In a broad sense, the proteome (i.e., the genome operating system) means all the proteins produced by a cell or tissue. Proteomics will contribute to bridge the gap between our understanding of genome sequence and cellular behavior. Proteomics offers an excellent way to study the reaction of the host and pathogen proteomes (i.e., genome-operating systems) during their complex biochemical cross talk.[20,21] Using the first generation proteomics approach, two-dimensional electrophoresis (2-DE), and mass spectrometry (MS), posttranslational modifications of host and pathogen proteins (such as phosphorylation, glycosylation, acetylation, and methylation) in reaction to their interaction can be detected. Such modifications are vital for the correct activity of numerous proteins and are being increasingly recognized as a major mechanism in cellular regulation. Although 2-DE offers a high-quality approach for the study of host and/or pathogen proteomes, during the post-genomic era several proteomics approaches (e.g., bottom-up; top-down) and quantitative proteomics strategies have been developed, which complement classical 2-DE (see Fig. 11.1).[17,22–26] Table 11.1 presents a comparison of the most popular proteomics tools.

3. Retrospective Analysis of Previous Proteomics Studies

The host–pathogen cross talks reflect the balance of host defenses and pathogen virulence mechanisms. Post-genomic technology promises to revolutionize many fields in biology by providing enormous amounts of genetic data from model and nonmodel organisms. Proteomics is a case point and promises to bridge the gap between our understanding of genome sequences and cellular behavior involved in host–pathogen interactions. Proteomics offers the possibility to characterize host–pathogen interactions from a global proteomic view. To date, most proteomics surveys on host–parasite interactions have focused on cataloguing protein content of pathogens and identifying virulence-associated proteins or proteomic alterations in host response to a pathogen. Also, many parasitologists and molecular biologists have used proteomics to find pathogen-specific host biomarkers for rapid pathogen detection and characterization of host–pathogen cross talks during the infection process. In this section, a synthetic retrospective of previous proteomics studies on host–pathogen interactions and some pitfalls of these surveys are presented.

Downloaded From Vet-eBooks.com

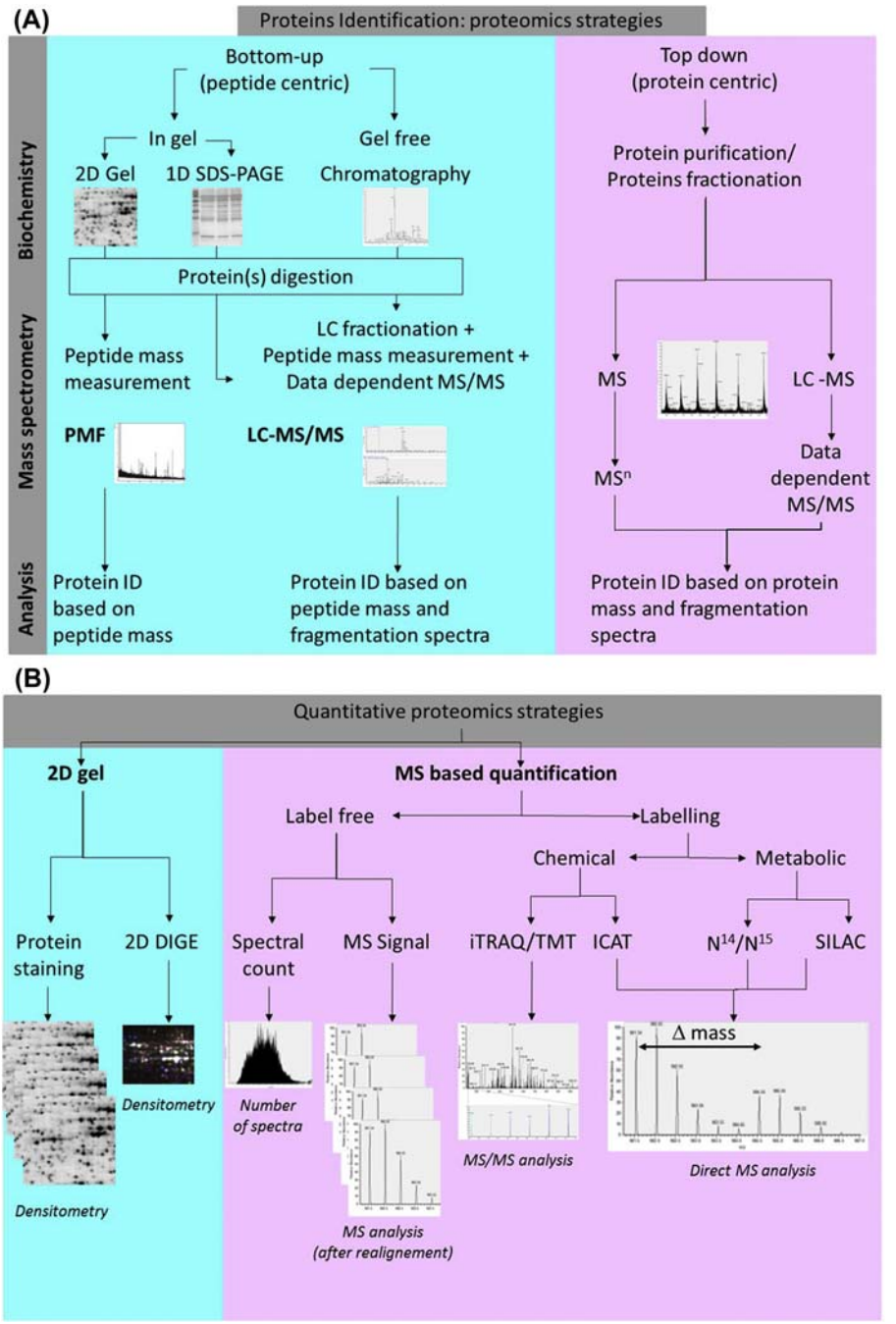


Figure 11.1 Deciphering of host−parasite cross talk with proteomics. Bottom-up and Top-down approaches (A); and quantitative proteomics strategies (B).

Downloaded From Vet-eBooks.com

Table 11.1 A Comparison of Proteomics Tools

Name of Technique	Separation	Quantification	Identification of Candidate Protein Spots	Hydrophobic Proteins	Requirement for Protein Identification	Potential for Discovering New Proteins	Detection of Specific Isoforms	Relative Assay Time	Cost to Acquire and to Use
2-DE	Electrophoresis: IEF PAGE	Densitometry of stains	Mass spectrometry (PMF; MS/MS)	Dependent on detergents used	No	Yes	Yes	Moderate	Cheap
2-DIGE	Electrophoresis: IEF PAGE	Densitometry of Cy3- and Cy5-labeled proteins normalize to Cy2	Mass spectrometry (PMF; MS/MS)	Dependent on detergents used	No	Yes	Yes	Moderate	Expensive
MuDPIT	LC−LC of peptides	None	Mass spectrometry (MS/MS)	Theoretically better than electrophoresis but not systematically examined	No	Yes	Yes	Rapid	Moderate
ICAT™	LC of peptides	Through use of heavy and light tags	Mass spectrometry (MS/MS)	No better than 2-DE	No	Yes	No	Rapid	Moderate
iTRAQ	LC of peptides	Labeling with isobaric mass tags	Mass spectrometry (MS/MS)	No better than 2-DE	No	Yes	Yes	Rapid	Expensive
SILAC	LC of peptides	Metabolic labeling with enriched stable isotope of amino acids ([$^{13}C_6^{15}N_4$]arginine and/or [$^{13}C_6^{15}N_2$] lysine).	Mass spectrometry (MS/MS)	No better than 2-DE	No	Yes	Yes	Rapid	Moderate

Continued

Downloaded From Vet-eBooks.com

Table 11.1 A Comparison of Proteomics Tools—cont'd

Name of Technique	Separation	Quantification	Identification of Candidate Protein Spots	Hydrophobic Proteins	Requirement for Protein Identification	Potential for Discovering New Proteins	Detection of Specific Isoforms	Relative Assay Time	Cost to Acquire and to Use
SELDI-TOF MS	Binding of proteins based on their chemical and physical characteristics	Comparison of MS peaks	Requires series of samples or coupling to second MS instrument	Moderate	No	Yes	No	Rapid	Expensive
Protein arrays	Antibody-based chips (binding to affinity reagent)	Densitometry of binding	Binding to particular affinity reagent	Unknown	Yes	No	Yes	Rapid	Cheap

2-DE, two-dimensional electrophoresis; *2-DIGE*, two-dimensional difference in gel electrophoresis; *ICAT*, isotope coded affinity tags; *iTRAQ*, isobaric tag for relative and absolute quantification; *LC*, liquid chromatography; *LC—LC*, tandem liquid chromatography; *MS/MS*, tandem mass spectrometry; *MuDPIT*, multidimensional protein identification technology; *PAGE*, polyacrylamide gel electrophoresis; *PMF*, peptide mass fingerprint; *SELDI-TOF MS*, spectrum enhanced laser desorption ionization-time of flight mass spectrometry; *SILAC*, stable isotope labeling by amino acids in cell culture.

Downloaded From Vet-eBooks.com

3.1 Deciphering of the Molecular Strategies Involved in Parasite Immune Evasion

To elude the vigilance of the immune system of a host, particularly mammals, a causative microorganism must actually act as a double agent. Indeed, the broad immunity has a natural or innate and adaptive component. Innate immunity constitutes the first antimicrobial defense and rapidly induces soluble mediators, such as complement, inflammatory cytokines, and chemokines, together with effector cells, such as macrophages and natural killers, in order to control or delay the spreading of the infectious agent. Then a specific response of adaptive immunity will take place to eliminate pathogens that would have survived innate immune response.[27] These immune selective pressures have conducted pathogens to develop mechanisms to modulate and alter host responses or to evade phagocytosis. As a result of these host−pathogen interactions, protein expression profiles of the host immune system (susceptibility/tolerance factors) and of the pathogen (virulence/pathogenicity factors) are mutually modified.[28−30]

Depending on the pathogen type (virus, bacteria, fungi, and unicellular or multicellular parasites), strategies of interactions will be different and the subversion of the host immune responses will exhibit specificities at the protein level (for reviews see Refs. 20,31,32). In fact, these molecular dialogues and conflicts can be seen as a chess game between the host immune cell populations and the pathogen populations, in which the pathogen plays with the whites (i.e., it starts the game). Because of differences in host−pathogen organisms' size and ratio, leading to size differences of respective proteomes, the pathogen proteome could be considered as overwhelmed by the host proteome during the interactions. But in terms of immune evasion, this is not limiting because the immune system works on a qualitative basis, which constitutes a second advantage for the pathogen that can induce large-scale damages with low amounts of molecules. By contrast, this represents not only one major limitation to characterize host−pathogen interactions, but also a challenging perspective for proteomics technology. This is why retrospectively proteomics studies were mainly conducted to evidence pathogenic virulence and pathogenicity factors.[33−43]

Independent of the proteomics workflow used for analysis, parasite immune evasion could be illustrated by at least three strategies that are commonly widespread among pathogens: (1) immune evasion based on antigenic variation, (2) inhibition of adaptive immunity activation systems, and (3) host mimicry. In African trypanosomes, the antigenic variation of the variant surface glycoprotein (VSG) constituting the surface coat of the parasite is well described.[44] But as in proteomics study, the parasite population, which has switched the VSG, is so poorly represented that it goes undetected, and therefore always keeps one step ahead of host immune responses. Also in trypanosomatids, *Leishmania amastigotes*, which establish within macrophage (a major immune effector cell), developed the ability to degrade class II major histocompatibility molecules to prevent Th1-type immunity to be induced.[45] Another protozoan parasite, *Toxoplasma gondii*, generates its parasitophorous vacuole with elements of the plasma membrane from the targeted host cells, thus using the host "self" to evade immune recognition.[46] These few examples actually perfectly illustrate how difficult it

Downloaded From Vet-eBooks.com

is to decipher, at the protein level during interactions, the pathogen molecular components involved in immune evasion. However, a new quantitative proteomics tools, the SILAC (stable isotope labeling by amino acids in cell culture) lately allowed the detection of 148 proteins of a microsporidian parasite during the kinetics of a host infection.[47] Among these proteins, many are involved in parasite proliferation, and an overrepresentation of putative secreted effectors proteins was observed. Finally, this SILAC survey also suggests that this microsporidia species could use a transposable element as a lure strategy to escape the host innate immune system. Advances in proteomics offer challenging perspectives to decipher the molecular war in host–pathogen interactions.[26]

3.2 Host Proteome Responses to Parasite Infection

While it seems obvious to say that when a pathogen will infect a host, the later will react by expressing molecules that can be characterized by clinical proteomics, it is surprising how few studies are devoted to this research. Yet the discovery of biomarkers signing an infected state from a healthy state is the heart of the Infectious Disease Research,[48,49] and expression proteomics has quickly developed to characterize the differential expression of proteins encoded by a particular gene and their posttranslational modifications in biological fluids and tissues.[50–52] In characterizing the host proteome responses to a pathogen infection, different levels of analysis have to be considered: soluble biomarkers expressed in biological fluids (e.g., serum, saliva, urine, and cerebrospinal fluid), tissue biomarkers indicative of an organ response and cellular biomarkers indicative of a cell-type response (e.g., immune cells).

Interestingly, the majority of the proteomics studies on host response to infection were performed on viral deregulation of host cells proteome ex vivo.[53–61] These works allowed to characterize at the molecular level the overall modifications in protein profiles of the target cells, and were of high interest to the better understanding of the pathogen influence on its host. In bacteria, studies have evaluated the mode of action of known toxins or bacterial components on host cells.[62,63] Concerning parasites, ex vivo experiments on host–parasite interactions have highlighted molecular details of manipulation strategies suffered by target cells during toxoplasmosis Chagas' disease or malaria.[64–66] Curiously, few works directly focused on the subversion of the immune system, mainly through monocyte/macrophage deregulation.[67,68]

As a paradox, the most striking studies on host proteome response to parasite infection were performed on arthropod (infectious diseases vectors)–parasite interactions. Probably because the parasite induced a strong phenotype modification,[69,70] particularly in the case of insect behavior manipulation.[71,72] Although few in number, taken together, these pioneering analyses of the response of the proteome of the host to a pathogen pave the way for the dynamic analysis of host–pathogen interactions. These approaches deserve to be strengthened and extended to all infectious diseases to increase and improve knowledge of the molecular dialogue and conflict that govern host–pathogen interactions.

On the other hand, the clinical aspect is important in infectious diseases, a number of studies have sought to characterize more comprehensively the proteome

Downloaded From Vet-eBooks.com

response of the host to infection in biological fluids, with a purpose diagnosis. One interesting pioneering study was performed in rabbits and allowed to detect intra-amniotic infection by proteomic-based amniotic fluid analysis.[73] For human diseases or those of livestock, the biological fluid, which should enable the detection of infection, linked to host proteome response, in the host serum. Several studies performed on this biological sample have allowed discriminating host-commensal from host−pathogen interactions in *Candida albicans*[74] and determining the immunome of pathogens.[75,76] Moreover, in African trypanosomiasis, proteomics analysis of the serum not only was indicative of the host response to infection, but also was promising for characterizing disease progression toward neurological disorder.[77,78] This illustrates how proteomics will help in considering at different analytical levels the host proteome response to a pathogen infection, with the prospect of benefits in improving diagnostics and therapeutics.

3.3 Biomarkers Linked to Infection Process by a Pathogen Using SELDI-TOF-MS Technology

High-throughput proteomic technology offers promise for the discovery of disease biomarkers and have extended our ability to unravel proteomes. In this section, we focus on the Surface-enhanced laser desorption time of flight mass spectrometry (SELDI-TOF-MS) technology. This MS-based method requires a minimal amount of sample for analysis and allows the rapid high-throughput analysis of complex protein samples.[79] SELDI-TOF-MS differs from conventional matrix−assisted laser desorption ionization (MALDI)-TOF-MS because the target surfaces, to which the proteins and matrices are applied to, are coated with various chemically active ProteinChip surfaces (ion exchange, immobilized metal affinity capture, and reverse phase arrays). Therefore, it is possible to fractionate proteins within a mixture, or particular classes of proteins, on the array surface prior to analysis. As with MALDI, different matrices can be used to facilitate the ionization and desorption of proteins from the SELDI array surface.[80]

This technology was initially applied to the discovery of early diagnostic or prognostic biomarkers of cancer.[81−83] Subsequently, this technology was used to discover fluid or tissue protein biomarkers for infectious diseases, such as HIV-1,[84−89] hepatitis B and C viruses,[66,90−93] severe acute respiratory syndrome[94] and BK virus,[95] African trypanosomiasis,[78,96] infection of Artemia by cestodes,[97] tuberculosis,[98] bacterial endocarditis,[99] and *Helicobacter pylori* infection.[100]

Certain individuals are resistant to HIV-1 infection, despite repeated exposure to the virus. The analysis of resistance to HIV infection is one of the research avenues, which has the hope of resulting in the development of a more effective treatment or a successful preventive vaccine against HIV infection. However, the molecular mechanism underlying resistance in repeatedly HIV-1-exposed, uninfected individuals (EU) is unclear. A complementary transcriptome and SELDI-TOF-MS analyses have been performed on peripheral blood T cells, plasma or serum from EU, their HIV-1-infected sexual partners, and healthy controls.[86] This study detected a specific

Downloaded From Vet-eBooks.com

biomarker associated with innate host resistance to HIV infection, as an 8.6-kDa A-SAA cleavage product.

In the same vein, understanding the virus−host interactions that lead to patients with acute hepatitis C virus (HCV) infection to viral clearance is a key toward the development of more effective treatment and prevention strategies. SELDI-TOF-MS technology has been used to compare, at a proteomic level, plasma samples, respectively, from donors who had resolved their HCV infection after seroconversion, from donors with chronic HCV infection, and from unexposed healthy donors.[92] A candidate marker of about 9.4 kDa was found to be higher in donors with HCV clearance than in donors with chronic infection. This biomarker was identified by nanoLC-Q-TOF-MS/MS as Apolipoprotein C-III and validated by Western Blot analysis. Among the most strongly upregulated genes in Dengue virus−infected *Aedes aegypti* salivary glands, one study identified a gene belonging to the cecropin family. The overexpression of this antimicrobial peptide was confirmed using the SELDI-TOF-MS technique.[101]

4. Toward New Conceptual Approaches to Decipher the Host−Parasite Interactions for Parasites With Simple or Complex Life Cycle

One main goal of "parasite-proteomics" surveys is to find proteins for use as pathogen-specific host biomarkers and to decipher the host−pathogen cross talks. Some papers emphasize that a significant number of surveys were done with a nonrigorous experimental design and without a conceptual approach to disentangle a general host proteome response from a specific host proteome response during the interaction with a pathogen.[12,20,30,43,102] A new attitude is essential to improve the reliability of proteomics data on host−pathogen interactions. Lately, some conceptual approaches have been proposed to researchers working on host−pathogen interactions to improve the reliability of "parasite-proteomics" results and to stimulate the creation of proteomic database with a holistic view of host−pathogen interactions. Thus, in this section, three new avenues to decipher host−pathogen interactions for any pathogen species (i.e., with simple or complex life cycle) are presented.

4.1 A Holistic Approach to Disentangle the Host and Parasite Genome Responses During Their Interactions

Some proteomics studies have shown common features in the innate response of plants, insects, and mammals.[103−106] The plant defense response is mediated by disease-resistance genes (R genes), which are abundant throughout the genome and confer resistance to many microorganisms, nematodes, and/or insects. R genes of several families of plants studied to date show homology with the *Drosophila* receptor Toll and the mammalian interleukin-1 receptor. In addition, plants, invertebrates, and vertebrates produce a class of peptides called "defensins," which are

Downloaded From Vet-eBooks.com

pathogen-inducible.[103] Some peptides and/or proteins used by phytophagous or animal parasites to modify the genome expression of their host share many structural and functional homologies. Thus, for example, phytoparasitic root-knot nematodes of the genus *Meloidogyne* secrete substances into their plant hosts in order to make a giant cell used as a feeding site.[107,108] A similar system is observed for the zooparasite, *Trichinella spiralis* (Stichosomida: Trichinellidae).[109] Furthermore, the injection of a peptide isolated from nematode secretions to either plant protoplasts or human cells enhances cell division.[110] The mechanism is not yet well known, but protein induction is considered as a strong possibility.

These days, many data are obtained by genomic and proteomics projects concerned with host−parasite interactions. Nevertheless, as mentioned earlier, generally little effort is made to elaborate such projects with respect to a holistic view of the goal to increase knowledge concerning immune responses of a host along with the biochemical cross talk between host and pathogen/parasite. Thus far, "parasito-proteomics" studies are in their infancy but have already led to new insights concerning molecular pathogenesis and microorganism identification.[111,112] However, many "parasito-proteomics" studies have been done with powerful tools, but without a conceptual approach to disentangle the host and parasite genome responses during their interactions.

Lately, a new holistic approach was proposed to parasitologists and molecular biologists based on evolutionary concepts of the immune response of a host to an invading parasite (for more details see Ref. 20). For instance, this new conceptual approach enables the classification of the host genomic response to infection by a parasite according to the immune mechanisms used (constitutive versus induced) and the degree of specificity. From an evolutionary-ecological point of view, host immune responses to a particular parasite can be plotted on a chart according to the immune mechanisms used (constitutive versus induced) and degree of specificity. The first axis of the defense chart refers to the immune mechanisms employed by the host with the two extreme cases: (1) a constitutive immune mechanism used by the host to rapidly impair the invasion by a parasite and (2) an induced immune mechanism, which has the advantage of avoiding a costly defense system, yet has the disadvantage that the parasite might escape host control.[15] The second axis of the defense chart refers to the degree of specificity of the host immune response.

Whatever the tactics used and the degree of specificity, the host genome ensures the adequate operation of the immune response via the proteome (genome operating system). For each immune tactic, many proteins are implicated. Consequently, any researcher in parasito-proteomics working with the immune defense chart will be able to categorize the host genome reaction for any given parasite at any given time. Also, for the pathogen, from an evolutionary-ecological point of view, parasite molecular strategies used to counteract host immune system can be plotted on a chart according to the infection mechanisms used (constitutive versus induced) and degree of specificity. This type of approach should be as much hypothesis generating for parasito-proteomics as for evolutionary ecology itself.

Lately, pioneer proteomics studies on parasite-induced alteration of host behavior (widespread transmission strategy among pathogens) have been carried out on six

Downloaded From Vet-eBooks.com

arthropod host−parasite associations: two orthoptera−hairworm associations, two insect vector−pathogen associations, and two gammarid−parasite associations.[113] These "parasito-proteomics" studies were based on the conceptual approach suggested by Biron et al.[20,21] Thus, in each study, many biological treatments have been effected to control the potential confusion resulting from proteins that are nonspecific to the manipulative process and to find the protein potentially linked with host behavioral changes. Also, for each study, to limit the possible effects of multiple infection and/or host sex-specific factors on the host proteome response, only monoinfected host males were used for the proteomics analysis. These "parasito-proteomics" surveys on the parasitic manipulation hypothesis showed that proteomic tools and the conceptual approach suggested by Biron et al.[20,21] are sensitive enough to disentangle host proteome alterations, and also the parasite proteome alterations linked to many factors, such as the circadian cycle, the parasitic status, parasitic emergence, the quality of a habitat, and the manipulative process.

4.2 Pathogeno-Proteomics: A New Avenue to Decipher Host−Vector−Pathogen Interactions

Relationships between pathogens and their hosts and vectors depend on a molecular dialogue tightly regulated. The reciprocal influence of a pathogen with its host or vector will affect the level of their genomes and their expression, respectively.[30] Variability and cross-regulation increase from genomic DNA (mutations, rearrangement, methylations, and so on) through RNA transcripts (initiation, splicing, maturation, editing, stability, and so on) to functional proteins (initiation, folding, posttranslational modifications, localization, function, and so on). Pathogeno-proteomics is a new approach to decipher host−vector−pathogen interactions, which integrates modifications at all analytical levels (genome, transcriptome, proteome: whole cell content, and secretome: naturally excreted−secreted molecules) through the analysis of their end-products' profile (Fig. 11.2). The concept is based on a management with drawers

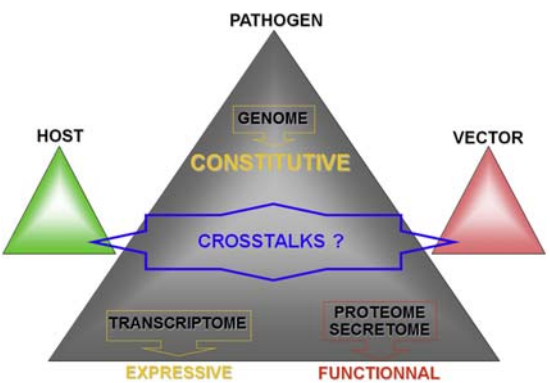


Figure 11.2 Pathogeno-proteomics: integrating analytical levels in host−vector−pathogen interactions.

Downloaded From Vet-eBooks.com

of the analytic workflow, from the determination of number of experimental treatments and design of the biological material preparation to the dedicated proteomics and bioinformatics tools needed to answer a research question in cell immunobiology (directly involved in host−pathogen interactions (Fig. 11.3)) but also in ecology and evolution, population's biology, and adaptive processes.[10,30,114] Moreover, it has been proved that the results of this type of integrated approach has a concrete impact on the discovery of the causes of infectious diseases, as well as on improving the diagnosis, vaccine development, and rational drug design.[115−117] Despite a theoretical aspect,[118] the pathogeno-proteomics concept brought new insights into important aspects of cell signaling[119] and molecular medicine.[120,121] As an example, proteomics and bioinformatics tools enable the formulation of relevant biological hypothesis on why part of the fungal population is killed while a significantly high percentage survives in *C. albicans*−macrophage interactions,[122] leading to addition of a specific database for studying *C. albicans*−host interactions.[123] Direct applications in terms of discovery of antifungal drug targets or design of new effective antibacterial vaccines become reality.[40,124] Other studies have also highlighted the pathogenic changes in the brain of SIV-infected monkeys,[125] adaptive metabolic changes in *Trypanosoma cruzi* and *Trypanosoma congolense*,[126,127] or molecular biomarkers of intestinal disorder induced by *H. pylori* or *Tritrichomonas muris*.[100,128] Subsequently, the use of model organisms interacting with infectious agent of medical importance emphasized the complexity and pathogen specificity of the worm's immune response.[129] Taken together, these examples demonstrate the potential of the concept of pathogeno-proteomics and promote this new research avenue.

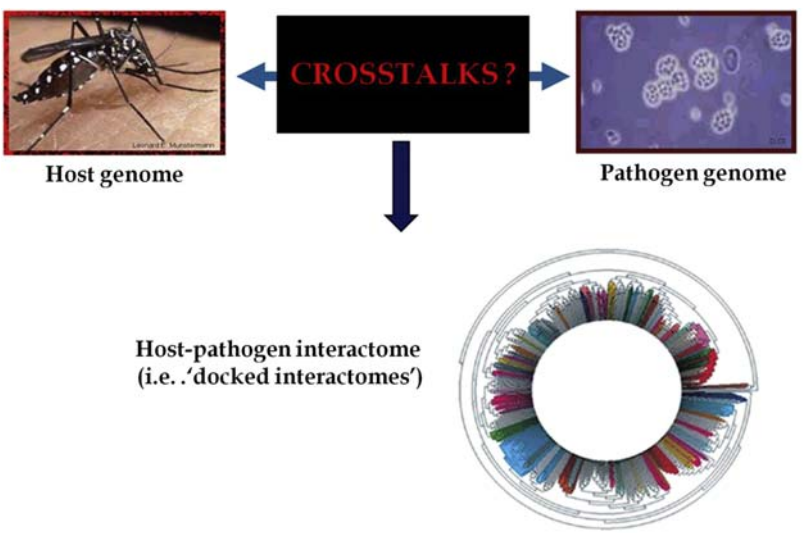


Figure 11.3 A new biological entity named host−pathogen interactome corresponding to complete set of protein−protein interactions existing between all the proteins of a host and a pathogen during their interaction.

Downloaded From Vet-eBooks.com

5. Population Proteomics: An Emerging Discipline to Study Host–Parasite Interactions

The host susceptibility to a pathogen and/or the pathogen virulence are often fluctuating within a host population even when infected hosts are collected in the same habitat and at the same time. This host phenotypic variability can be caused by three factors: (1) host genotype and/or pathogen genotype, (2) different environmental experiences (e.g., habitat fragmented in microclimates), and (3) host coinfection by pathogens (i.e., competition or mutualism among coinfecting pathogens within hosts). What are the host–pathogen cross talks at individual and population scales in a habitat? Is it possible to detect and to decipher the host proteome variability within a habitat for the molecular mechanisms and for the protein networks involved in the host–pathogen interactions? In this section, a new emerging discipline in proteomics, the population proteomics, and its prospects are presented with results of some pioneer studies on this topic, especially in human population proteomics.

5.1 Prospects With Population Proteomics for Any Living Organisms

One limiting factor for the first generation of proteomics tools (e.g., 2-DE) is the amount of proteins required to study the host and/or pathogen proteome expression(s) during their interactions. Most surveys in "parasite-proteomics" were done by pooling many individuals for any treatment (e.g., infected and noninfected hosts) required to answer a query. Thus, with this kind of experimental protocol, no data can be acquired on the interindividual variation in expression of host and pathogen proteomes during their cross talk. New proteomics tools and methods have been developed as 2D-LC/MS that can permit to study the interindividual variation of molecular cross talk in host–pathogen associations.[130–132]

At the beginning of the century, Dobrin Nedelkov proposed a new scientific field in proteomics: the population proteomics.[130] Population proteomics was defined as the study of protein diversity in human populations, or more specifically, targeted investigation of human proteins across and within populations to define and understand protein diversity with the main aim to discover disease-specific protein modulations.[133] Biron et al.[114] have proposed to broaden the "population proteomics" concept to all living organisms with the aims to complement the population genetics and to offer a new avenue to decipher the cross talk diversity involved in trophic interactions in a habitat since the execution of the genetic plan is carried out by the activities of proteins and natural selection acts at the protein level.[10,134]

The apparent separation between genomics and proteomics that leads to different perspective on the same ecological reality is a fundamental limitation that needs to be overcome if complex processes, such as adaptation, pathogen virulence, and host susceptibility, are to be understood. Population proteomics coupled with population genetics has a great potential to resolve issues specific to the ecology, the evolution of natural populations, the dynamic of host susceptibility to pathogens, the evolution of pathogen virulence, and the range of host genotypes that can be infected with a

Downloaded From Vet-eBooks.com

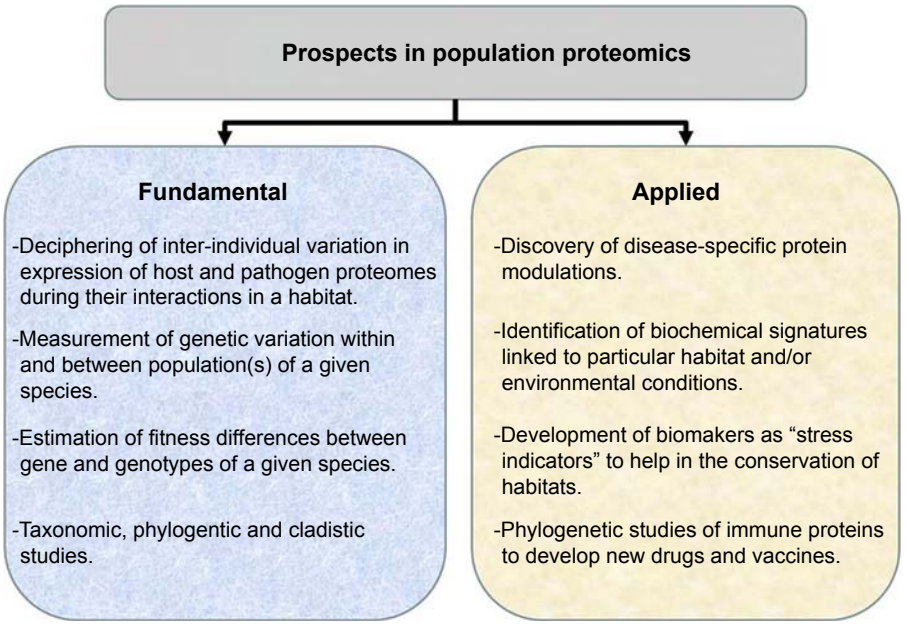


Figure 11.4 Potential of population proteomics as an emerging discipline in proteomics.

given pathogen genotype in host−parasite interactions. Some perspectives for the population proteomics are resumed in Fig. 11.4. Even if we are yet far from this "promised land," a better understanding of the information contained in proteomics markers should permit an impressive amount of information to be gathered on the past as well as current environmental conditions experienced by a given population of a species, something that could be summarized as "show me your proteome and I will tell you who you are, where you are from, and where you should go from here."

Lately, pioneer surveys on population proteomics have been carried out with classical proteomic tools (i.e., 2-DE and MS) (1) to determine the genetic variability between species and between populations of a given species,[135−137] (2) to identify biochemical signatures linked to particular habitat and/or environmental conditions,[138,139] and (3) for phylogenetic studies.[130,140] Nedelkov et al.[141,142] have investigated the human plasma proteins diversity by using approaches similar to enzyme-linked immunosorbent assay but utilizing MS as method of detection.[133] These pioneer results should help not only to discover disease-specific protein modulations but also to find pathogen-specific protein biomarkers. The next subsection presents in detail the Nedelkov' results on protein diversity in human populations.

5.2 Human Population Proteomics

Human population proteomics deciphers protein diversity in human populations. In a broader term, human population proteomics can be compared to human population genomics, where individuals are interrogated with the aim of cataloguing common

Downloaded From Vet-eBooks.com

genetic variants and determining how they are distributed among people within populations and among populations in different parts of the world.[143–145] Although human population proteomics cannot (yet) claim such outreach and goals, it has the potential to become an important proteomics subdiscipline as the tools and approaches that enable it become more embraced and practiced.

Human population proteomics does not engage the study of entire proteomes because it is very likely that, for a specific cell or tissue proteome, there is no definitive set and number of proteins that is common to all within a group or a larger population. Instead, human population proteomics focuses on interrogation of a selected number of proteins but from a large number of individuals, to delineate the distribution of specific protein modifications within these subpopulations. Hence, targeted protein analysis approaches utilizing MS as detection method are employed. MS measures a unique feature of each fully expressed protein—its molecular mass. Changes in the protein structure resulting from structural modifications are reflected in its molecular mass and can be detected via MS, without a priori knowledge of the modification. The MS methods utilized in human population proteomics must be capable of analyzing hundreds, if not thousands of samples per day, with high reproducibility and sensitivity. Hence, top-down MS approaches utilizing affinity ligands are the most likely methods of choice for population proteomics.[144] Surface-immobilized ligands can be utilized to affinity-retrieve a protein of interest from a biological sample, after which the protein (with or without the affinity ligand) is introduced into a mass spectrometer. One of the first affinity MS methods developed was mass spectrometric immunoassay (MSIA).[146] The approach combines targeted protein affinity-extraction with rigorous characterization using MALDI-TOF MS (Fig. 11.5). Protein(s) are extracted from a biological sample with the help of affinity pipettes derivatized with polyclonal antibodies. The proteins are eluted from the affinity pipettes with a MALDI matrix, and are MS-analyzed. Enzymatic digestion, if needed, is performed on the MALDI target itself. Specificity and sensitivity, as in traditional immunoassays, are dictated by the affinity-capture reagents—the antibodies.

However, a second measure of specificity is incorporated in the resulting mass spectra, wherein each protein registers at specific m/z value. During data analysis, the major signal in the mass spectrum that corresponds to the targeted protein is initially evaluated; it should be within a reasonable range (e.g., error of measurement of $<0.05\%$) from the value of the empirically calculated mass obtained from the sequence of the protein deposited in the Swiss-Prot databank. Once this mass value is confirmed (or observed to be shifted), the presence of protein modifications is noted by the appearance of other signals in the mass spectra (usually in the vicinity of the native protein peaks), or by mass shifts of the major protein signal. Modifications can be tentatively assigned by accurate measurement of the observed mass shifts (from the wild-type protein signals and/or in silico calculated mass) and knowledge of the protein sequence and possible modifications. The identity of the modifications is then verified using proteolytic digestion and mass mapping approaches in combination with high-performance MS.

In an initial study of human protein diversity using MS methods of detection, 25 plasma proteins from a cohort of 96 healthy individuals were investigated via MS

Downloaded From Vet-eBooks.com

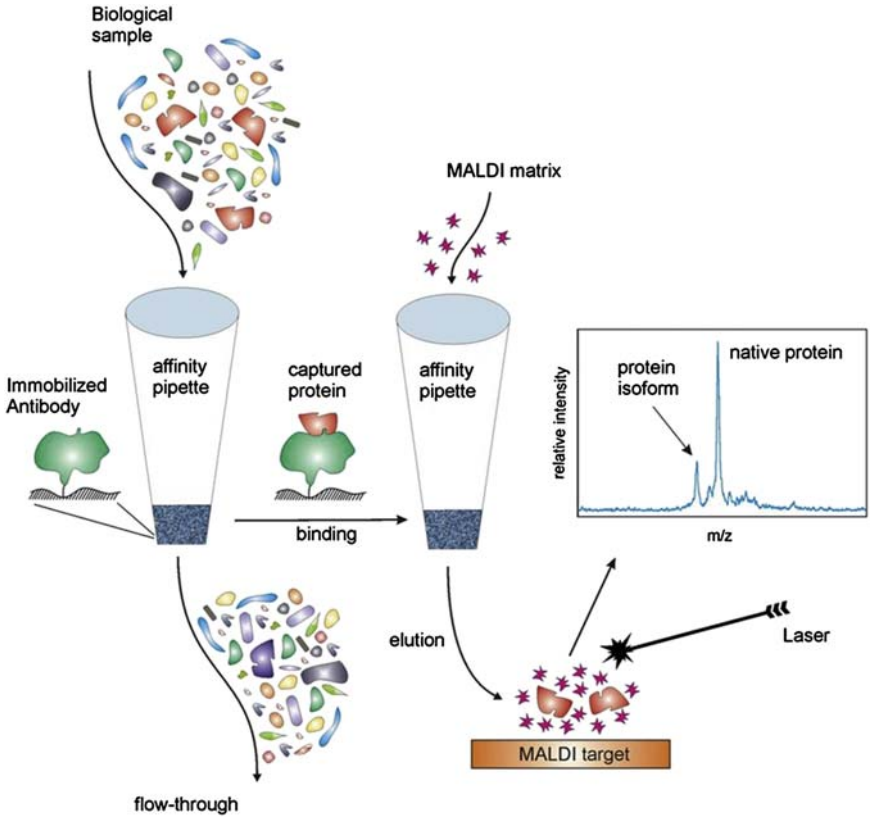


Figure 11.5 Schematics of the mass spectrometric immunoassay (MSIA) approach.

immunoassays.[147] The protocol and an example of the data generated for one of the proteins, transthyretin (TTR), are outlined in Fig. 11.6. The TTR MSIA assays were performed in parallel on the 96 human plasma samples using affinity pipettes derivatized with anti-TTR antibody. Following MS analysis, data matrix containing all tentatively assigned modifications was assembled. Then, peptide-mapping experiments were performed on selected number of samples to identify the specific modifications and finalize the modifications database. The data for all 25 proteins is presented in Fig. 11.7, which lists the modifications observed for 18 of the 25 proteins studied (modifications were not observed for 7 proteins), and shows the frequency of each modification in the 96 samples cohort. A total of 53 protein variants were observed for these 18 proteins, stemming from posttranslational modifications and point mutations. The largest number of posttranslationally modified protein variants was found to be C- or N-terminal truncated protein isoforms. Deglycosylation, oxidation, and cysteinylation were also observed among several of the proteins. Among the point mutations detected for four of the proteins, notable was the high incidence of point mutations for apolipoprotein E and TTR, which is consistent with genomic studies that have found these proteins to be highly polymorphic. The overall frequency of

Downloaded From Vet-eBooks.com

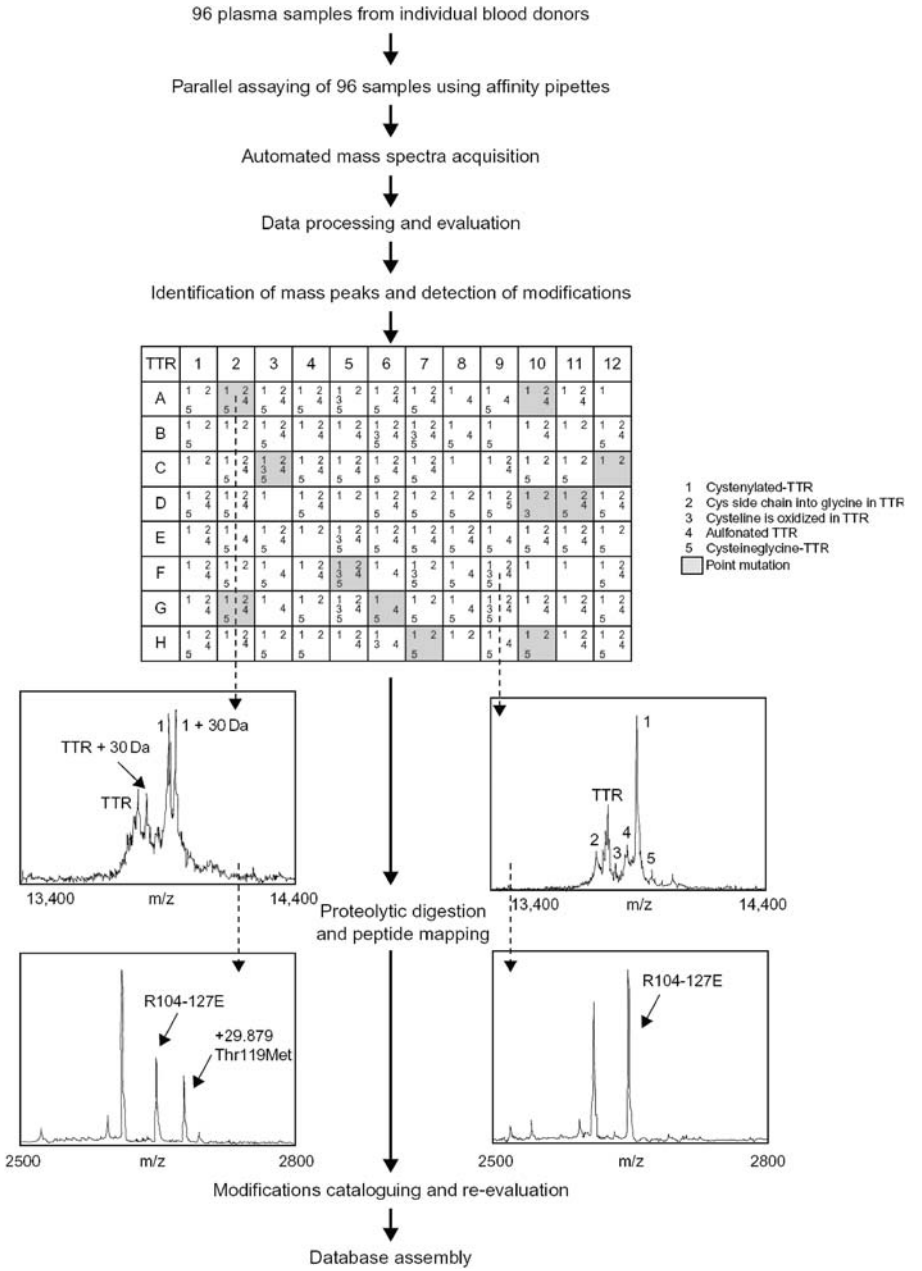


Figure 11.6 An outline of a population proteomics approach using TTR as an example. *m/z*, mass-to-charge ratio; *TTR*, transthyretin.

Downloaded From Vet-eBooks.com

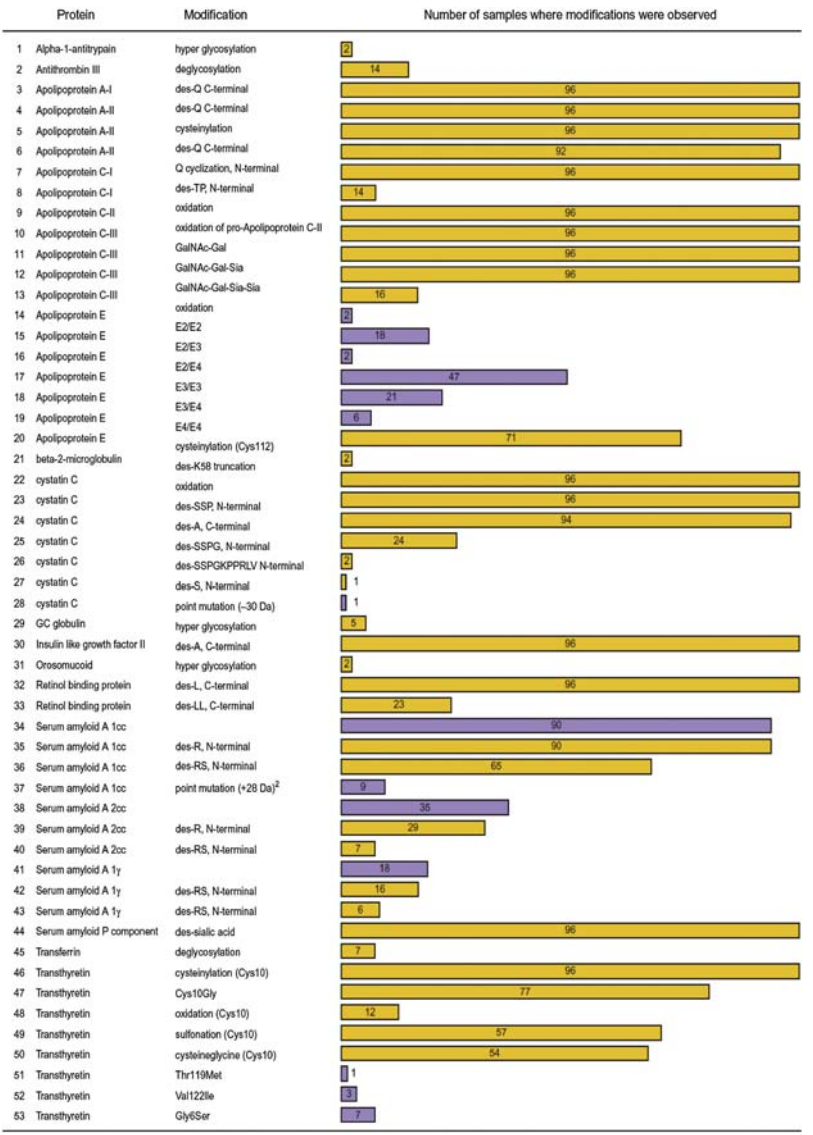


Figure 11.7 Modifications observed in 18 of the 25 proteins analyzed from 96 human plasma samples.

the modifications in the 96 samples cohort was wide ranged. Fourteen modifications were observed in all 96 samples, suggesting that they must be regarded as wild-type protein forms. Others, such as most of the point mutations, were present in only few of the samples. Overall, 23 of the modifications were observed in more than 65% of the samples, and 20 in less than 15% of the 96 samples analyzed. Upon further data analysis, and taking into the consideration the gender, age, and ethnicity of the

Downloaded From Vet-eBooks.com

individuals who provided the samples, it was determined that the Gly6Ser mutation in TTR was detected only in individuals of Caucasian origin, which is consistent with existing knowledge about the occurrence of this common non-amyloidogenic population polymorphism in Caucasians.[148,149] Another correlation was observed in regard to interprotein variations in specific individuals: all seven individuals for which carbohydrate-deficient transferrin was detected were also characterized with deglycosylated antithrombin III.

Following this small-scale protein diversity study, a second study of human protein diversity was carried out wherein the number of samples was greatly expanded in order to get an accurate view of the distribution of some of the protein modifications in the general population.[142] Thousand individuals from four geographical regions in the United States (California, Florida, Tennessee, and Texas) were selected, and the protein modifications for beta-2-microglobulin (b2m), cystatin C (cysC), retinol-binding protein (RBP), transferrin (TRFE), and TTR were delineated (in the 96 samples study, these five proteins accounted for 19 of the 53 protein variants observed). The results of the study are summarized in Fig. 11.8, which lists the protein modifications observed and the frequency of each in the 1000 samples cohort. A total of 27 protein modifications (20 posttranslational modifications and 7 point mutations) were detected, with various frequencies in the cohort of samples. Variants resulting from oxidation were observed most frequently, along with single amino acid truncations. Least frequent were variants arising from point mutations and extensive sequence truncations. In total, 6 modifications were observed with high frequency (present in $>80\%$ of the samples), 5 were of medium frequency (20−50% of the samples), and 16 were low-frequency modifications observed in $<7\%$ of the samples. Nine of the low-frequency modifications were not observed in the 96 individuals study. Thus, by increasing the size of the population, it became possible to detect these low-occurrence protein modifications. When the frequencies of the modifications in the two studies were compared, an excellent correlation was obtained. For example, in both cohorts about 7% of the individuals were characterized with carbohydrate-deficient transferrin. Upon further data analysis based on the gender, age, and geographical origin of the individuals who provided the samples, it was determined that the samples obtained from California contained significantly less protein modifications than the samples obtained from Florida, Tennessee, and Texas, even though the samples from all four states were collected in the same way within a 3-month window in the spring of 2005, and stored under identical conditions until analysis. Correlations were also made in regard to the gender distribution of two protein modifications. Carbohydrate-deficient transferrin was observed in about 1% of the females and about 10% of the males in the 1000 cohort. Carbohydrate-deficient transferrin is an FDA-approved clinical biomarker for alcoholism, and this gender correlation can partially be explained by the higher prevalence of alcohol dependence in males than in females. The second gender correlation was related to cystatin C: all 10 of the cystatin C point mutations were found in males.

Two conclusions can be made from these two systematic studies of protein modifications and variants. First, MS is capable of detecting structural protein modifications, and, when coupled to immunoaffinity separations, it can be employed in a

Downloaded From Vet-eBooks.com

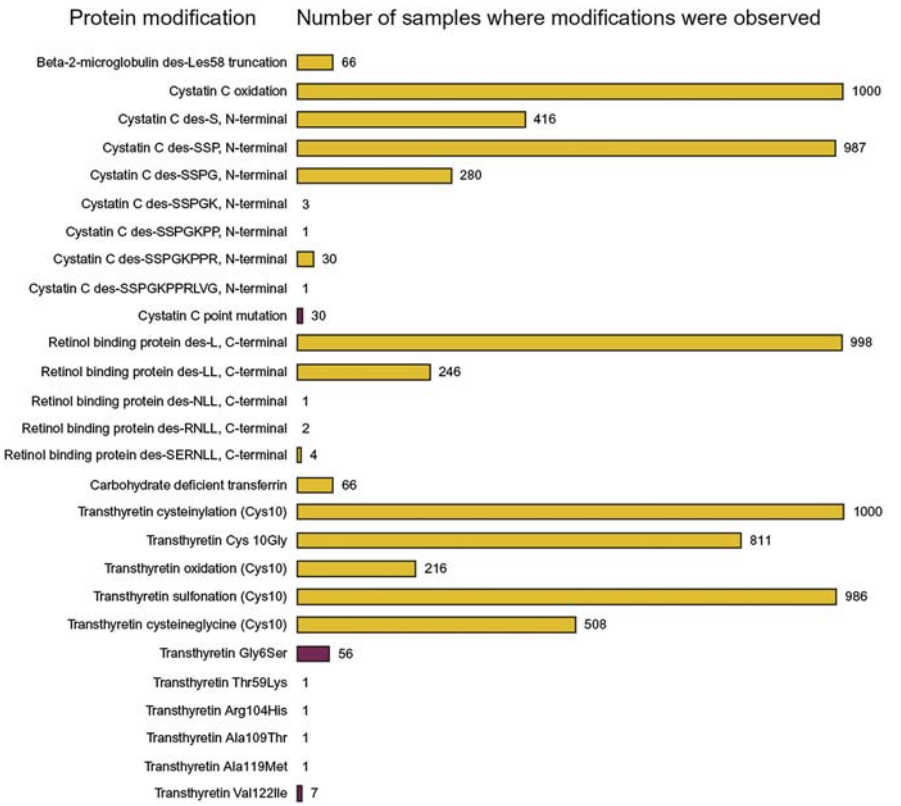


Figure 11.8 Modifications observed for five proteins studied from 1000 human plasma samples.

high-throughput systematic study of human protein diversity. Second, the human protein diversity is far more complex than the variation observed at the genetic level. While it might be premature to declare the human proteins variation "the next big thing," it is reasonable to predict that assessing human proteome variations among and within populations will be a paramount effort that can facilitate biomarker discovery. Such endeavor would represent a paradigm shift in proteomics with significant clinical and diagnostic implications, as protein variations, quantitative and qualitative, begin to be associated with specific diseases.

6. Conclusion

From the dawn of human evolution to the influenza and HIV/AIDS pandemics of the 20th and early 21st centuries, infectious diseases have continued to emerge and re-emerge with great ferocity and, by so doing, seriously affect populations as well as challenge our abilities to fight the responsible agents. Over the past decade, strains of many common pathogens have continued to develop resistance to the drugs that

Downloaded From Vet-eBooks.com

once were effective against them. In the battle against pathogens, humankind has created new mega-technologies, such as massive sequencing, proteomics, and bioinformatics, but without conceptual approaches based on the evolutionary concepts. Parasite genome sequences do not of themselves provide a full explanation of the biology of an organism and on the molecular war involved in host−pathogen associations. Since the 1990s, proteomic tools have been successfully employed in a large number of studies to find and identify proteins involved in biological phenomena, for example, immunity, host−parasite interactions, and so on. Even so, many studies have, as outlined earlier, revealed pitfalls in the approaches used. Thus, whatever the new technological advancements, it is apparent that parasitologists and molecular biologists should attempt to improve their experimental design. This new attitude will surely improve the reliability of the data deriving from proteomics studies and will open the way for an enhanced comprehension of many biological mechanisms. In this chapter, new ways based on evolutionary concepts are suggested to enable further elucidation of the molecular complexities of host−pathogen genome interactions. These new ways could help to increase the knowledge about the molecular war involved in host−pathogen associations taking into account the environmental factors.

References

1. Achtman M, Morelli G, Zhu P, Wirth T, Diehl I, Kusecek B, et al. Microevolution and history of the plague bacillus, *Yersinia pestis*. *Proc Natl Acad Sci USA* 2004;**101**: 17837−42.
2. Watts S. *Epidemics and history: disease, power and Imperialism*. New Haven, CT: Yale University Press; 1997.
3. Hochstrasser DF. Proteome in perspective. *Clin Chem Lab Med* 1998;**36**:825−36.
4. Degrave WM, Melville S, Ivens A, Aslett M. Parasite genome initiatives. *Int J Parasitol* 2001;**3**:532−6.
5. Ouma JH, Vennervald BJ, Butterworth AE. Morbidity in schistosomiasis: an update. *Trends Parasitol* 2001;**17**:117−8.
6. Ryan ET. Malaria: epidemiology, pathogenesis, diagnosis, prevention, and treatment-an update. *Curr Clin Top Infect Dis* 2001;**21**:83−113.
7. Guzman MG, Kouri G. Dengue: an update. *Lancet Inf Dis* 2002;**2**:33−42.
8. Gelfand JA, Callahan MV. Babesiosis: an update on epidemiology and treatment. *Curr Infect Dis Rep* 2003;**5**:53−8.
9. WHO. *Global Health Observatory (GHO) Data*. http://www.who.int/gho/malaria/epidemic/deaths/en/; 2016.
10. Karr TL. Application of proteomics to ecology and population biology. *Heredity* 2008; **100**:200−6.
11. Barret J, Jefferies JR, Brophy PM. Parasite proteomics. *Parasitol Today* 2000;**16**:400−3.
12. Ashton PD, Curwen RS, Wilson RA. Linking proteome and genome: how to identify parasite proteins. *Trends Parasitol* 2001;**17**:198−202.
13. Fell DA. Beyond genomics. *Trends Genet* 2001;**17**:680−2.
14. Fields S. Proteomics in genomeland. *Science* 2001;**291**:1221−4.
15. Schmid-Hempel P, Ebert D. On the evolutionary ecology of specific immune defence. *Trends Ecol Evol* 2003;**18**:27−32.

Downloaded From Vet-eBooks.com

16. Anderson L, Seilhaver J. A comparison of selected mRNA and protein abundance in human liver. *Electrophoresis* 1997;**18**:533−7.
17. Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* 1999;**19**:1720−30.
18. Maniatis T, Tasic B. Alternative pre-mRNA splicing and proteome expression in metazoans. *Nature* 2002;**418**:236−43.
19. Anderson NG, Anderson NL. Twenty years of two-dimensional electrophoresis: past, present and future. *Electrophoresis* 1996;**17**:443−53.
20. Biron DG, Moura H, Marché L, Hughes AL, Thomas F. Towards a new conceptual approach to 'Parasitoproteomics'. *Trends Parasitol* 2005;**2005**(21):162−8.
21. Biron DG, Joly C, Galéotti N, Ponton F, Marché L. The proteomics: a new prospect for studying parasitic manipulation. *Behav Process* 2005;**68**:249−53.
22. Fung ET, Thulasiraman V, Weinberger SR, Dalmasso EA. Protein biochips for differential profiling. *Curr Opin Biotechnol* 2001;**12**:65−9.
23. Lopez MF, Pluskal MG. Protein micro- and macroarrays: digitizing the proteome. *J Chromatogr B* 2003;**787**:19−27.
24. Wu CC, MacCoss MJ, Howell KE, Yates III JR. A method for the comprehensive proteomics analysis of membrane proteins. *Nat Biotechnol* 2003;**21**:532−8.
25. Bischoff R, Luider TM. Methodological advances in the discovery of protein and peptide disease markers. *J Chromatogr B* 2004;**803**:27−40.
26. Chetouhi C, Panek J, Bonhomme L, El Alaoui H, Texier C, Langin T, et al. Cross-talk in host−parasite associations: what do past and recent proteomics approaches tell us? *Inf Genet Evol* 2015;**33**:84−94.
27. Roitt IM, Delves PJ. *Essential immunology*. Blackwell Publishing, Inc; 2001.
28. Zhang CG, Chromy BA, McCutchen-Maloney SL. Host-pathogen interactions: a proteomic view. *Expert Rev Proteomics* 2005;**2**:187−282.
29. Coiras M, Camafeita E, López-Huertas MR, Calvo E, López JA, Alcamí J. Application of proteomics technology for analyzing the interactions between host cells and intracellular infectious agents. *Proteomics* 2008;**8**:852−73.
30. Holzmuller P, Grébaut P, Brizard JP, Berthier D, Chantal I, Bossard G, et al. "Pathogenoproteomics": towards a new approach of host-vector-pathogen interactions. *Ann NY Acad Sci* 2008;**1149**:66−70.
31. Walduck A, Rudel T, Meyer TF. Proteomic and gene profiling approaches to study host responses to bacterial infection. *Curr Opin Microbiol* 2004;**7**:33−8.
32. Viswanathan K, Früh K. Viral proteomics: global evaluation of viruses and their interaction with the host. *Expert Rev Proteomics* 2007;**4**:815−29.
33. Ouellette M, Olivier M, Sato S, Papadopoulou B. Studies on the parasite *Leishmania* in the post-genomic era. *Med Sci* 2003;**19**:900−9.
34. Texier C, Brosson D, El Alaoui H, Méténier G, Vivarès CP. Post- genomics of microsporidia, with emphasis on a model of minimal eukaryotic proteome: a review. *Folia Parasitol* 2005;**52**:15−22.
35. Van Hellemond JJ, van Balkom BW, Tielens AG. Schistosome biology and proteomics: progress and challenges. *Exp Parasitol* 2007;**117**:267−74.
36. Liu N, Song W, Wang P, Lee K, Chan W, Chen H, et al. Proteomics analysis of differential expression of cellular proteins in response to avian H9N2 virus infection in human cells. *Proteomics* 2008;**8**:1851−8.
37. Bird DM, Opperman CH. The secret(ion) life of worms. *Genome Biol* 2009;**10**:205.
38. Steuart RF. Proteomic analysis of Giardia: studies from the pre- and post-genomic era. *Exp Parasitol* 2010;**124**:26−30.

Downloaded From Vet-eBooks.com

39. Weiss LM, Fiser A, Angeletti RH, Kim K. *Toxoplasma gondii* proteomics. *Expert Rev Proteomics* 2009;**6**:303−13.
40. Jagusztyn-Krynicka EK, Dadlez M, Grabowska A, Roszczenko P. Proteomic technology in the design of new effective antibacterial vaccines. *Expert Rev Proteomics* 2009;**6**: 315−30.
41. Premsler T, Zahedi RP, Lewandrowski U, Sickmann A. Recent advances in yeast organelle and membrane proteomics. *Proteomics* 2009;**9**:4731−43.
42. Bhavsar AP, Auweter SD, Finlay BB. Proteomics as a probe of microbial pathogenesis and its molecular boundaries. *Future Microbiol* 2010;**5**:253−65.
43. Holzmuller P, Grébaut P, Cuny G, Biron DG. Tsetse flies, trypanosomes, humans and animals: what is proteomics revealing about their crosstalks? *Expert Rev Proteomics* 2010; **7**:113−26.
44. Morrison LJ, Marcello L, McCulloch R. Antigenic variation in the African trypanosome: molecular mechanisms and phenotypic complexity. *Cell Microbiol* 2009;**11**:1724−34.
45. Antoine JC, Prina E, Courret N, Lang T. *Leishmania* spp.: on the interactions they establish with antigen-presenting cells of their mammalian hosts. *Adv Parasitol* 2004;**58**:1−68.
46. Plattner F, Soldati-Favre D. Hijacking of host cellular functions by the Apicomplexa. *Annu Rev Microbiol* 2008;**62**:471−87.
47. Panek J, El Alaoui H, Mone A, Urbach S, Demettre E, Texier C, et al. Hijacking of host cellular functions by an intracellular parasite, the microsporidian Anncaliia algerae. *PLoS One* 2014;**9**:e100791.
48. te Pas MF, Claes F. Functional genomics and proteomics for infectious diseases in the post-genomics era. *Lancet* 2004;**363**:1337.
49. Azad NS, Rasool N, Annunziata CM, Minasian L, Whiteley G, Kohn EC. Proteomics in clinical trials and practice: present uses and future promise. *Mol Cell Proteomics* 2006;**5**: 1819−29.
50. Fournier I, Wisztorski M, Salzet M. Tissue imaging using MALDI-MS: a new frontier of histopathology proteomics. *Expert Rev Proteomics* 2008;**5**:413−24.
51. Hood BL, Malehorn DE, Conrads TP, Bigbee WL. Serum proteomics using mass spectrometry. *Methods Mol Biol* 2009;**520**:107−28.
52. Wilm M. Quantitative proteomics in biological research. *Proteomics* 2009;**9**:4590−605.
53. Zheng X, Hong L, Shi L, Guo J, Sun Z, Zhou J. Proteomics analysis of host cells infected with infectious bursal disease virus. *Mol Cell Proteomics* 2008;**7**:612−25.
54. Liu HC, Hicks J, Yoo D. Proteomic dissection of viral pathogenesis. *Dev Biol* 2008;**132**: 43−53.
55. Lee SR, Nanduri B, Pharr GT, Stokes JV, Pinchuk LM. Bovine viral diarrhea virus infection affects the expression of proteins related to professional antigen presentation in bovine monocytes. *Biochim Biophys Acta* 2009;**1794**:14−22.
56. Sun J, Jiang Y, Shi Z, Yan Y, Guo H, He F, et al. Proteomic alteration of PK-15 cells after infection by classical swine fever virus. *J Proteome Res* 2008;**7**:5263−9.
57. Pastorino B, Boucomont-Chapeaublanc E, Peyrefitte CN, Belghazi M, Fusaï T, Rogier C, et al. Identification of cellular proteome modifications in response to West Nile virus infection. *Mol Cell Proteomics* 2009;**8**:1623−37.
58. Vester D, Rapp E, Gade D, Genzel Y, Reichl U. Quantitative analysis of cellular proteome alterations in human influenza A virus-infected mammalian cell lines. *Proteomics* 2009;**9**: 3316−27.
59. Antrobus R, Grant K, Gangadharan B, Chittenden D, Everett RD, Zitzmann N, et al. Proteomic analysis of cells in the early stages of herpes simplex virus type-1 infection reveals widespread changes in the host cell proteome. *Proteomics* 2009; **9**:3913−27.

Downloaded From Vet-eBooks.com

60. Zhang X, Zhou J, Wu Y, Zheng X, Ma G, Wang Z, et al. Differential proteome analysis of host cells infected with porcine circovirus type 2. *J Proteome Res* 2009;**8**:5111−9.
61. Zhang L, Jia X, Zhang X, Sun J, Peng X, Qi T, et al. Proteomic analysis of PBMCs: characterization of potential HIV-associated proteins. *Proteome Sci* 2010;**8**:12.
62. Kuhn JF, Hoerth P, Hoehn ST, Preckel T, Tomer KB. Proteomics study of anthrax lethal toxin-treated murine macrophages. *Electrophoresis* 2006;**27**:1584−97.
63. Shui W, Gilmore SA, Sheu L, Liu J, Keasling JD, Bertozzi CR. Quantitative proteomic profiling of host-pathogen interactions: the macrophage response to *Mycobacterium tuberculosis* lipids. *J Proteome Res* 2009;**8**:282−9.
64. Teixeira PC, Iwai LK, Kuramoto AC, Honorato R, Fiorelli A, Stolf N, et al. Proteomic inventory of myocardial proteins from patients with chronic Chagas' cardiomyopathy. *Braz J Med Biol Res* 2006;**39**:1549−62.
65. Nelson MM, Jones AR, Carmen JC, Sinai AP, Burchmore R, Wastling JM. Modulation of the host cell proteome by the intracellular apicomplexan parasite Toxoplasma gondii. *Infect Immun* 2008;**76**:828−44.
66. Wu Y, Nelson MM, Quaile A, Xia D, Wastling JM, Craig A. Identification of phosphorylated proteins in erythrocytes infected by the human malaria parasite Plasmodium falciparum. *Malar J* 2009;**8**:105.
67. Oura CA, McKellar S, Swan DG, Okan E, Shiels BR. Infection of bovine cells by the protozoan parasite *Theileria annulata* modulates expression of the ISGylation system. *Cell Microbiol* 2006;**8**:276−88.
68. Fischer J, West J, Agochukwu N, Suire C, Hale-Donze H. Induction of host chemotactic response by *Encephalitozoon* spp. *Infect Immun* 2007;**75**:1619−25.
69. Biron DG, Marché L, Ponton F, Loxdale HD, Galéotti N, Renault L, et al. Behavioural manipulation in a grasshopper harbouring hairworm: a proteomics approach. *Proc R Soc Lond B* 2005;**272**:2117−26.
70. Rachinsky A, Guerrero FD, Scoles GA. Differential protein expression in ovaries of uninfected and Babesia-infected southern cattle ticks, *Rhipicephalus* (Boophilus) microplus. *Insect Biochem Mol Biol* 2007;**37**:1291−308.
71. Lefèvre T, Thomas F, Ravel S, Patrel D, Renault L, Le Bourligu L, et al. *Trypanosoma brucei brucei* induces alteration in the head proteome of the tsetse fly vector *Glossina palpalis gambiensis*. *Insect Mol Biol* 2007;**16**:651−60.
72. Lefevre T, Thomas F, Schwartz A, Levashina E, Blandin S, Brizard JP, et al. Malaria *Plasmodium* agent induces alteration in the head proteome of their *Anopheles* mosquito host. *Proteomics* 2007;**7**:1908−15.
73. Klein LL, Freitag BC, Gibbs RS, Reddy AP, Nagalla SR, Gravett MG. Detection of intra-amniotic infection in a rabbit model by proteomics-based amniotic fluid analysis. *Am J Obstet Gynecol* 2005;**193**:1302−6.
74. Pitarch A, Nombela C, Gil C. Proteomic profiling of serologic response to *Candida albicans* during host-commensal and host-pathogen interactions. *Methods Mol Biol* 2009; **470**:369−411.
75. Sakolvaree Y, Maneewatch S, Jiemsup S, Klaysing B, Tongtawe P, Srimanote P, et al. Proteome and immunome of pathogenic *Leptospira* spp. revealed by 2DE and 2DE-immunoblotting with immune serum. *Asian Pac J Allergy Immunol* 2007;**25**: 53−73.
76. Ju JW, Joo HN, Lee MR, Cho SH, Cheun HI, Kim JY, et al. Identification of a serodiagnostic antigen, legumain, by immunoproteomic analysis of excretory-secretory products of *Clonorchis sinensis* adult worms. *Proteomics* 2009;**9**:3066−78.
77. Papadopoulos MC, Abel PM, Agranoff D, Stich A, Tarelli E, Bell BA, et al. A novel and accurate diagnostic test for human African trypanosomiasis. *Lancet* 2004;**363**:1358−63.

Downloaded From Vet-eBooks.com

78. Agranoff D, Stich A, Abel P, Krishna S. Proteomic fingerprinting for the diagnosis of human African trypanosomiasis. *Trends Parasitol* 2005;**21**:154−7.
79. De Bock M, de Seny D, Meuwis MA, Chapelle JP, Louis E, Malaise M, et al. Challenges for biomarker discovery in body fluids using SELDI-TOF-MS. *J Biomed Biotechnol* 2010: 906082.
80. Merchant M, Weinberger SR. Recent advancements in surface-enhanced laser desorption/ ionization-time of flight-mass spectrometry. *Electrophoresis* 2000;**21**:1164−77.
81. Petricoin EF, Liotta LA. SELDI-TOF-based serum proteomic pattern diagnostics for early detection of cancer. *Curr Opin Biotechnol* 2004;**15**:24−30.
82. Yang SY, Xiao XY, Zhang WG, Zhang LJ, Zhang W, Zhou B, et al. Application of serum SELDI proteomic patterns in diagnosis of lung cancer. *BMC Cancer* 2005;**20**(5):83.
83. Xiao Z, Prieto D, Conrads TP, Veenstra TD, Issaq HJ. Proteomic patterns: their potential for disease diagnosis. *Mol Cell Endocrinol* 2005;**230**:95−106.
84. Luo X, Carlson KA, Wojna V, Mayo R, Biskup TM, Stoner J, et al. Macrophage proteomic fingerprinting predicts HIV-1-associated cognitive impairment. *Neurology* 2003;**60**: 1931−7.
85. Sun B, Rempel HC, Pulliam L. Loss of macrophage-secreted lysozyme in HIV-1-associated dementia detected by SELDI-TOF mass spectrometry. *AIDS* 2004;**18**: 1009−12.
86. Missé D, Yssel H, Trabattoni D, Oblet C, Lo Caputo S, Mazzotta F, et al. IL-22 participates in an innate anti-HIV 1 host-resistance network through acute-phase protein induction. *J Immunol* 2007;**178**:407−15.
87. Luciano-Montalvo C, Ciborowski P, Duan F, Gendelman HE, Meléndez LM. Proteomic analyses associate cystatin B with restricted HIV-1 replication in placental macrophages. *Placenta* 2008;**29**:1016−23.
88. Toro-Nieves DM, Rodriguez Y, Plaud M, Ciborowski P, Duan F, Pérez Laspiur J, et al. Proteomic analyses of monocyte derived macrophages infected with human immunodeficiency virus type 1 primary isolates from Hispanic women with and without cognitive impairment. *J Neurovirol* 2009;**15**:36−50.
89. Wiederin J, Rozek W, Duan F, Ciborowski P. Biomarkers of HIV-1 associated dementia: proteomic investigation of sera. *Proteome Sci* 2008;**17**:7−8.
90. Poon TC, Hui AY, Chan HL, Ang IL, Chow SM, Wong N, et al. Prediction of liver fibrosis and cirrhosis in chronic hepatitis B infection by serum proteomic fingerprinting: a pilot study. *Clin Chem* 2005;**51**:328−35.
91. Kanmura S, Uto H, Kusumoto K, Ishida Y, Hasuike S, Nagata K, et al. Early diagnostic potential for hepatocellular carcinoma using the SELDI ProteinChip system. *Hepatology* 2007;**45**:948−56.
92. Molina S, Misse D, Roche S, Badiou S, Cristol JP, Bonfils C, et al. Identification of apolipoprotein C-III as a potential plasmatic biomarker associated with the resolution of hepatitis C virus infection. *Proteomics Clin Appl* 2008;**2**:751−61.
93. Fujita N, Sugimoto R, Motonishi S, Tomosugi N, Tanaka H, Takeo M, et al. Patients with chronic hepatitis C achieving a sustained virological response to peginterferon and ribavirin therapy recover from impaired hepcidin secretion. *J Hepatol* 2008;**49**: 702−10.
94. Pang RT, Poon TC, Chan KC, Lee NL, Chiu RW, Tong YK, et al. Serum proteomic fingerprints of adult patients with severe acute respiratory syndrome. *Clin Chem* 2006;**52**: 421−9.
95. Jahnukainen T, Malehorn D, Sun M, Lyons-Weiler J, Bigbee W, Gupta G, et al. Proteomic analysis of urine in kidney transplant patients with BK virus nephropathy. *J Am Soc Nephrol* 2006;**17**:3248−56.

Downloaded From Vet-eBooks.com

96. Stiles JK, Whittaker J, Sarfo BY, Thompson WE, Powell MD, Bond VC. Trypanosome apoptotic factor mediates apoptosis in human brain vascular endothelial cells. *Mol Biochem Parasitol* 2004;**133**:229–40.
97. Sánchez MI, Thomas F, Perrot-Minnot MJ, Biron DG, Bertrand-Michel J, Missé D. Neurological and physiological disorders in Artemia harboring manipulative cestodes. *J Parasitol* 2009;**95**:20–4.
98. Liu Q, Chen X, Hu C, Zhang R, Yue J, Wu G, et al. Serum protein profiling of smear-positive and smear-negative pulmonary tuberculosis using SELDI-TOF mass spectrometry. *Lung* 2010;**188**:15–23.
99. Fenollar F, Goncalves A, Esterni B, Azza S, Habib G, Borg JP, et al. A serum protein signature with high diagnostic value in bacterial endocarditis: results from a study based on surface-enhanced laser desorption/ionization time-of-flight mass spectrometry. *J Infect Dis* 2006;**194**:1356–66.
100. Wu MS, Chow LP, Lin JT, Chiou SH. Proteomic identification of biomarkers related to *Helicobacter pylori*-associated gastroduodenal disease: challenges and opportunities. *J Gastroenterol Hepatol* 2008;**23**:1657–61.
101. Luplertlop N, Surasombatpattana P, Patramool S, Dumas E, Wasinpiyamongkol L, Saune L, et al. Induction of a peptide with activity against a broad spectrum of pathogens in the *Aedes aegypti* salivary gland, following infection with Dengue virus. *PLoS Pathog* 2011;**7**(1):e1001252.
102. Tastet C, Bossis M, Gauthier JP, Renault L, Mugniéry D. *Meloidogyne chitwoodi* and *M. fallax* protein variation assessed by two-dimensional electrophoregram computed analysis. *Nematology* 1999;**1**:201–14.
103. Broekaert WF, Terras FRG, Cammue BPA, Osborn RW. Plants defensins: novel antimicrobial peptides as components of the host defense system. *Plant Physiol* 1995;**108**:1353–8.
104. Rock FL, Hardiman G, Timans JC, Kastelein RA, Bazan JF. A family of human receptors structurally related to *Drosophilla* Toll. *Proc Natl Acad Sci USA* 1998;**95**:588–93.
105. Cao H, Baldini RL, Rahme LG. Common mechanism for pathogens of plant and animals. *Annu Rev Phytopath* 2001;**39**:259–84.
106. Taylor JE, Hatcher PE, Paul ND. Crosstalk between plant responses to pathogens and herbivores: a view from the outside in. *J Exp Bot* 2003;**55**:159–68.
107. Abad P, Favery B, Ross MN, Castagnone-Sereno P. Root-knot nematode parasitism an host response: molecular basis of a sophisticated interaction. *Mol Plant Pathol* 2003;**4**: 217–24.
108. Doyle EA, Lambert KN. *Meloidogyne javanica* chorismate mutase 1 alters plant cell development. *Mol Plant Microbe Interact* 2003;**16**:123–31.
109. Jasmer DP. *Trichinella spiralis* infected skeletal muscle cells arrest in G2/M and cease muscle gene expression. *J Cell Biol* 1993;**121**:785–93.
110. Goverse A, De Engler JA, Verhees J, Van der Krol S, Helder JH, Gheysen G. Cell cycle activation by plant parasitic nematodes. *Plant Mol Biol* 2000;**43**:747–76.
111. Moura H, Ospina M, Woolfitt AR, Barr JR, Visvesvara GS. Analysis of four human microsporidian isolates by MALDI-TOF mass spectrometry. *J Eukaryot Microbiol* 2003; **50**:156–63.
112. Biron DG, Agnew P, Marché L, Renault L, Sidobre C, Michalakis Y. Proteome of *Aedes aegypti* larvae in response to infection by the intracellular parasite *Vavraia culicis*. *Int J Parasitol* 2005;**35**:1385–97.
113. Lefèvre T, Adamo SA, Biron DG, Missé D, Hughes D, Thomas F. Invasion of the body snatchers: the diversity and evolution of manipulative strategies in host–parasite interactions. In: Webster JP, Rollinson D, Hay SI, editors. *Advances in parasitology*, vol. 68. London: Academic Press; 2009. p. 46–84.

Downloaded From Vet-eBooks.com

114. Biron DG, Loxdale HD, Ponton F, Moura H, Marché L, Brugidou C, et al. Population proteomics: an emerging discipline to study metapopulation ecology. *Proteomics* 2006;**6**: 1712−5.
115. Doytchinova IA, Taylor P, Flower DR. Proteomics in vaccinology and immunobiology: an informatics perspective of the immunone. *J Biomed Biotechnol* 2003;**2003**:267−90.
116. Bansal AK. Bioinformatics in microbial biotechnology−a mini review. *Microb Cell Fact* 2005;**4**:19.
117. Chautard E, Thierry-Mieg N, Ricard-Blum S. Interaction networks: from protein functions to drug discovery. A review. *Pathol Biol* 2009;**57**:324−33.
118. Kint G, Fierro C, Marchal K, Vanderleyden J, De Keersmaecker SC. Integration of 'omics' data: does it lead to new insights into host-microbe interactions? *Future Microbiol* 2010;**5**: 313−28.
119. Kleppe R, Kjarland E, Selheim F. Proteomic and computational methods in systems modeling of cellular signaling. *Curr Pharm Biotechnol* 2006;**7**:135−45.
120. Ahram M, Petricoin EF. Proteomics discovery of disease biomarkers. *Biomark Insights* 2008;**23**:325−33.
121. Ostrowski J, Wyrwicz LS. Integrating genomics, proteomics and bioinformatics in translational studies of molecular medicine. *Expert Rev Mol Diagn* 2009;**6**:623−30.
122. Diez-Orejas R, Fernández-Arenas E. *Candida albicans*-macrophage interactions: genomic and proteomic insights. *Future Microbiol* 2008;**3**:661−81.
123. Vialás V, Nogales-Cadenas R, Nombela C, Pascual-Montano A, Gil C. Proteopathogen a protein database for studying *Candida albicans*—host interaction. *Proteomics* 2009;**9**: 4664−8.
124. Tournu H, Serneels J, Van Dijck P. Fungal pathogens research: novel and improved molecular approaches for the discovery of antifungal drug targets. *Curr Drug Targets* 2005;**6**:909−22.
125. Pendyala G, Trauger SA, Kalisiak E, Ellis RJ, Siuzdak G, Fox HS. Cerebrospinal fluid proteomics reveals potential pathogenic changes in the brains of SIV-infected monkeys. *J Proteome Res* 2009;**5**:2253−60.
126. Grébaut P, Chuchana P, Brizard JP, Demettre E, Seveno M, Bossard G, et al. Identification of total and differentially expressed excreted-secreted proteins from *Trypanosoma congolense strains exhibiting* different virulence and pathogenicity. *Int J Parasitol* 2009;**10**: 1137−50.
127. Roberts SB, Robichaux JL, Chavali AK, Manque PA, Lee V, Lara AM, et al. Proteomic and network analysis characterize stage-specific metabolism in *Trypanosoma cruzi*. *BMC Syst Biol* 2009;**16**(3):52.
128. Kashiwagi A, Kurosaki H, Luo H, Yamamoto H, Oshimura M, Shibahara T. Effects of *Tritrichomonas muris* on the mouse intestine: a proteomic analysis. *Exp Anim* 2009;**58**: 537−42.
129. Bogaerts A, Beets I, Temmerman L, Schoofs L, Verleyen P. Proteome changes of *Caenorhabditis elegans* upon a *Staphylococcus aureus* infection. *Biol Direct* 2010; **17**(5):11.
130. LaCount DJ, Vignali M, Chettier R, Phansalkar A, Bell R, Hesselberth JR, et al. A protein interaction network of the malaria parasite Plasmodium falciparum. *Nature* 2005;**438**: 103−7.
131. Navas A, Albar JP. Application of proteomics in phylogenetic and evolutionary studies. *Proteomics* 2004;**4**:299−302.
132. Predel R, Wegener C, Russell WK, Tichy SE, Russell DH, Nachman RJ. Peptidomics of CNS-associated neurohemal systems of adult *Drosophila melanogaster*: a mass spectrometric survey of peptides from individuals' flies. *J Comp Neurol* 2004;**474**:379−92.

Downloaded From Vet-eBooks.com

133. Brand S, Hahner D, Ketterlinus R. Protein profiling and identification in complex biological samples using LC-MALDI. *Drug Plus Int* September 2005:6−8.
134. Nedelkov D. Population proteomics: investigation of protein diversity in human populations. *Proteomics* 2008;**8**:779−86.
135. Cieslak A, Ribera I. Aplicaciones de proteómica en ecología y evolución. *Ecosistemas* 2009;**18**:34−43.
136. Chevalier F, Martin O, Rofidal V, Devauchelle AD, Barteau S, Sommerer N, et al. Proteomic investigation of natural variation between Arabidopsis ecotypes. *Proteomics* 2004; **4**:1372−81.
137. Diz AP, Skibinski DOF. Evolution of 2-DE protein patterns in a mussel hybrid zone. *Proteomics* 2007;**7**:2111−20.
138. Valcu CM, Lalanne C, Muller-Starck G, Plomion C, Schlink K. Protein polymorphism between 2 *Picea abies* populations revealed by 2-dimensional gel electrophoresis and tandem mass spectrometry. *Heredity* 2008;**99**:364−75.
139. Thiellement H, Bahrman N, Damerval C, Plomion C, Rossignol M, Santoni V, et al. Proteomics for genetic and physiological studies in plants. *Electrophoresis* 1999;**20**: 2013−26.
140. Pedersen KS, Codrea MC, Vermeulen CJ, Loeschcke V, Bendixen E. Proteomic characterization of a temperature-sensitive conditional lethal in *Drosophila melanogaster*. *Heredity* 2010;**104**:125−34.
141. Dorus S, Busby SA, Gerike U, Shabanowitz J, Hunt DF, Karr TL. Genomic and functional evolution of the *Drosophila melanogaster* sperm proteome. *Nat Genet* 2006;**38**:1440−5.
142. Nedelkov D, Kiernan UA, Niederkofler EE, Tubbs KA, Nelson RW. Investigating human plasma proteins diversity. *Proc Natl Acad Sci USA* 2005;**102**:10852−7.
143. Nedelkov D, Phillips DA, Tubbs KA, Nelson RW. Investigation of human protein variants and their frequency in the general population. *Mol Cell Proteomics* 2007;**6**:1183−7.
144. Nedelkov D, Tubbs KA, Niederkofler EE, Kiernan UA, Nelson RW. High- throughput comprehensive analysis of human plasma proteins: a step toward population proteomics. *Anal Chem* 2004;**76**:1733−7.
145. Nedelkov D. Mass spectrometry-based immunoassays for the next phase of clinical applications. *Expert Rev Proteomics* 2006;**3**:631−40.
146. Nedelkov D, Kiernan UA, Niederkofler EE, Tubbs KA, Nelson RW. Population proteomics: the concept, attributes, and potential for cancer biomarker research. *Mol Cell Proteomics* 2006;**5**:1811−8.
147. Nelson RW, Krone JR, Bieber AL, Williams P. Mass-spectrometric immunoassay. *Anal Chem* 1995;**67**:1153−8.
148. Nedelkov D. Population proteomics: addressing protein diversity in humans. *Expert Rev Proteomics* 2005;**2**:315−24.
149. Connors LH, Lim A, Prokaeva T, Roskens VA, Costello CE. Tabulation of human transthyretin (TTR) variants. *Amyloid* 2003;**10**:160−84.

Downloaded From Vet-eBooks.com

This page intentionally left blank

Downloaded From Vet-eBooks.com

The Evolution of Antibiotic Resistance

12

F. González-Candelas[1,2], I. Comas[1,2,3], J.L. Martínez[4], J.C. Galán[2,5], F. Baquero[2,5]

[1]FISABIO/CSISP-UV/Instituto Cavanilles, Valencia, Spain; [2]CIBER en Epidemiología y Salud Pública, Madrid, Spain; [3]IBV-CSIC, Valencia, Spain; [4]CNB-CSIC, Madrid, Spain; [5]IRYCIS-Hospital Ramón y Cajal, Madrid, Spain

1. Introduction

Antibiotic resistance refers to the property of bacteria that prevents the inhibition of their growth by antimicrobial agents used in the clinical setting. The problem is dramatic in some countries[1] and especially worrying in highly pathogenic species, such as *Mycobacterium tuberculosis*,[2] methicillin-resistant *Staphylococcus aureus*,[3] or *Klebsiella pneumoniae*.[4] Antibiotic resistance represents one of the best examples of natural selection in action; and also one of the major hurdles in humankind's fight against infectious diseases. Here, we consider antibiotic resistance from a dual perspective, evolutionary and clinical, and hope to contribute to a better understanding of the principles and processes that result its emergence and to suggest strategies to prevent, or at least delay, its spread. Human actions may not prevent evolution but we can try to drive it through less-damaging pathways.

Antibiotic resistance can be "natural," when all the strains of the same bacterial species are resistant to a particular drug (intrinsic resistance), or "acquired," when there are susceptible and resistant strains in the same species, with resistant strains having evolved from susceptible ones by selection after mutation or horizontal gene transfer (HGT) events. In addition, there are situations, such as biofilm growth, stationary phase, or the presence of specific inducers of resistance, in which bacteria can present phenotypic, noninheritable resistance that does not involve genetic changes.

As the activity of antibiotics depends on their concentration, susceptibility or resistance is defined based on MIC (minimal inhibitory concentration, the lowest amount of antibiotic that inhibits visible bacterial growth) values. Clinical break points are based on the likelihood of therapeutic success and consider as resistant a microorganism if its MIC is higher than the concentration associated with favorable clinical outcomes. "Natural," in contrast to clinical, break points consider a bacterial organism as "resistant" if its MIC is significantly higher than the modal MIC of a collection of strains of the same bacterial species, thus considering resistance as an "abnormally higher" MIC. This definition encompasses all potential mechanisms of resistance acting at the population level, but requires the analysis of a large number of isolates and does not

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00012-3
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

consider hypersusceptible mutants, whose identification is the cornerstone for defining the intrinsic resistome. A final, *operational* definition of resistance derives from the comparison of a wild-type and a derived strain, either containing a mutation or a heterologous gene. If the antibiotic susceptibilities of wild-type and mutant strains are different, the mutated (or acquired) gene is involved in antibiotic resistance. This definition can be applied to the functional mining of bacterial genomes and metagenomes.[5]

Antibiotic resistance and MIC values have been consistently rising for many bacterial species, even nonpathogenic ones, since the start of the industrial production of antimicrobial agents. Considering that most antibiotics act at concentrations close to 1 mg/mL, the total annual production of antibiotics is enough to cover the entire surface of the Earth with inhibitory concentrations; in other words, likely microbial populations part of these alterations are predictable, such as extended antibiotic resistance, but unpredictable effects are most likely to occur, including changes in the interactions among microbes or with multicellular organisms that influence basic cycles in the biosphere.[6,7]

The use of antibiotics in human health and other areas converges to a single, cooperative effect, changing bacterial ecology not only in different environments but also in the *common* environment. The main problem is the existing connectivity between all environments, human, farming, and agricultural, so that the antibiotic-derived effects, including selection and spread of resistance, in one of them have consequences in all the others.

The connection between different environments occurs essentially in two ways. First, the spread and migration of biological units, from genes to bacterial communities, play a major role in antibiotic resistance. Second, the dispersal of antimicrobial agents, which results in the production of selective mixed gradients and stressor effects, and in an acceleration of microbial evolutionary rates. The combination of migration of antibiotics and antibiotic-resistant biological units results in evolutionary activating interactions that occur in four main genetic reactors: (1) the intestinal microbiota of humans and animals; (2) the highly antibiotic-exposed areas with high rates of bacterial transmission, such as hospitals (particularly new-born wards and intensive-care units), (3) waste-water, effluents, and sewage treatment plants, and (4) soil, sediments, and surface and ground waters,[8] all of which contribute to the escalation of the emergence and spread of antimicrobial resistance.

The most evident threat of antimicrobial resistance for humankind is the failure of therapy against infectious diseases. The decrease in the incidence of infectious diseases in the Western world started in the beginning of the 19th century, by reasons related to social progress, better nutrition, and housing and hygienic procedures, but in the absence of antibiotics. The discovery and subsequent industrial production of antimicrobials between 1935 and 1960 was followed by a further reduction in the morbidity and mortality of infections, particularly the more severe ones, and has contributed to the increase in the expected duration of lifetime of human populations. At the same time, antibiotics facilitated the progress of Medicine at large, allowing interventions (complex surgery, intensive-care units, immunosuppressive and anticancer

Downloaded From Vet-eBooks.com

chemotherapy, or transplantation) that expose the impaired host to both pathogenic and opportunistic bacterial infections.

If antibiotic resistance was surpassing a threshold-limit, the consequences on the current standards of hospital-based medicine (including long-term care facilities for the elderly) could become severely compromised. With the emergence of multiresistant Gram-positive organisms, such as methicillin-resistant *S. aureus* (MRSA) or *Enterococcus faecium*, or Gram-negatives, such as pan-resistant *Escherichia coli*, *K. pneumoniae*, or *Pseudomonas aeruginosa* producing extended-spectrum β-lactamases (ESBL) and carbapenemases are currently very close to such a threshold. A transient equilibrium was reached during decades 1950−1980 because resistance was countered by the continuous discovery of novel antimicrobial agents active on resistant strains. Unfortunately, during the last quarter of the 20th century, no significant advances occurred in this field as a result of the interest of a number of pharmaceutical companies in investing more in chronic, noncurable diseases. Currently, resistance to the newest antibiotics continues evolving mostly on the bases of the old genetic mobile structures (plasmids, transposons, integrons) that became prevalent by the selective effect of the old antibiotics. The effect that the anthropogenic release of antibiotics has already caused on the genetic structure of bacterial populations is probably irreversible and will influence the evolutionary future of microbes on Earth.

Cleaning nature of this resistance gene pool is impossible. The best we can do is trying to control the emergence, selection, and spread of antibiotic-resistance genes in bacterial organisms interacting with humans, animals, or plants. The classical strategies for controlling the emergence of resistance are based on the reduction of chronic antibiotic−promoted bacterial mutagenic−stress associated with low dosages, the use of combinations of drugs, early intensive therapy, maintaining low bacterial density, and the surveillance of hypermutable organisms and the suppression of phenotypic resistance. A number of these strategies have been explored by population and mathematical modeling.[9,10] Controlling the selection of antibiotic resistance is a major practical goal, which can be addressed again by the development of novel antiinfective drugs and the appropriate use of antibiotics, avoiding low dosages able to select low-level mutations serving as stepping-stones for high-level resistance.

Avoiding the emergence of resistance in the individual patient has minimal effects at the community level. The efficacy of classical ways of controlling selection and spread is inversely proportional to the density and penetration of resistant organisms and their mobile genetic elements in particular environments. Measures that might be successful in the early stages of resistance development, or in settings with low rates of antibiotic resistance, are worthless in areas where resistance is already well established. Even in areas with low antibiotic resistance, such as Sweden, studies have shown that a 2-year discontinuation in the use of trimethoprim did not reduce significantly the rates of resistance to this compound.[11] This was probably due to the dispersion of trimethoprim-resistance genes in a multiplicity of bacterial organisms and mobile genetic elements frequently harboring other resistance determinants, thus assuring coselection of *dfr* genes with other resistance genes.

Some regions of the world are densely polluted with antibiotic resistance. In a global world, sooner or later, resistance originated in these "source of resistance" areas

Downloaded From Vet-eBooks.com

will invade still clean environments. Resistant organisms are constantly diluted and potentially extinguish in competition with constant immigration of susceptible bacteria in local environments, but such a trend might collapse by the increase of resistant populations. Moreover, the success of resistant organisms will contribute to the constant accumulation in the bacterial world of genetic platforms and vehicles able to efficiently recruit and spread novel resistance genes. Antibiotic resistance increases bacterial evolvability; resistance calls for more resistance, in a phenomenon described as "genetic capitalism."[12] In other words, resistance might be reversible when rare; if frequent, reversibility is not to be expected.

2. Mechanisms and Sources of Antibiotic Resistance

To produce an effect in a bacterial cell, an antibiotic has to cross different envelopes, occasionally be activated by bacterial enzymes, and reach its target at a high-enough concentration to allow a successful interaction and the inhibition of bacterial growth or killing (Fig. 12.1A). Resistance can be achieved either if the antibiotic concentration reaching the target is too low or if the interaction between the antibiotic and the target is not efficient enough to produce the inhibition of bacterial growth. This includes intrinsic and acquired resistance.

The most classical mechanisms of intrinsic resistance are the absence of the target and a reduced permeability to a given antibiotic. These are passive systems of

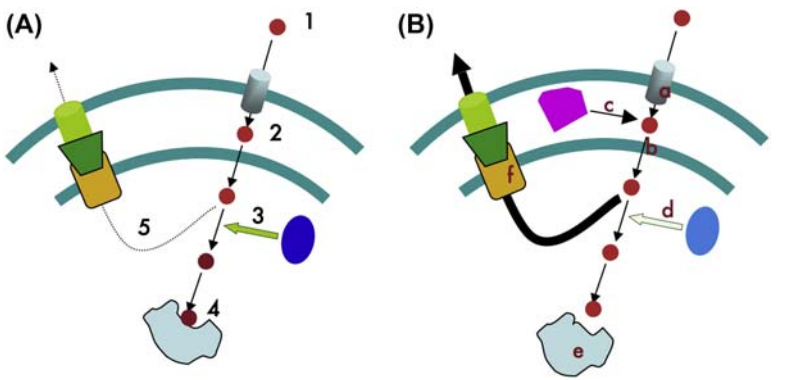


Figure 12.1 Basic mechanisms of antibiotic action and resistance. In order to inhibit bacterial growth, an antibiotic requires to successfully interact with its target at concentrations high enough for inhibiting its activity. For this (Panel A), the antibiotic (1) requires to traverse cellular envelopes (2), in some occasions to be activated by an intracellular enzyme (3) and reach its target (4). The activity of constitutively expressed MDR efflux pumps (5) can decrease the intracellular concentration of the antibiotic. As shown in Panel B, resistance is achieved by interfering with this pathway, either by changes that impede the entrance of the antibiotic (a, b), by the activity of antibiotic-inactivating enzymes (c) or overexpression of MDR efflux pumps (f) that reduce the effective intracellular concentration, or by mutations in the enzyme that activates the preantibiotic (d) or in its target (e), which preclude target/antibiotic interactions.

Downloaded From Vet-eBooks.com

resistance. However, bacterial populations also present active mechanisms of resistance based on the detoxification of the antibiotic. These include chromosomally encoded antibiotic inactivating enzymes and multidrug-resistance (MDR) efflux pumps. The analyses of comprehensive libraries of mutants, mainly from *E. coli*[13] and *P. aeruginosa*[14–16] have demonstrated that several genes participate in the intrinsic resistance phenotype of a bacterial species. This suggests that intrinsic resistance is not just a consequence of adaptation to the presence of a given antibiotic, but rather a phenotypic consequence of the general physiological characteristics of each species.

Contrary to intrinsic resistance, an ancient phenotype of bacterial populations, acquired resistance is the consequence of adaptive evolution to recent selective pressures exerted by the extended use of antibiotics.[7] Resistance can be achieved by avoiding the activity of the antibiotic at two main levels: changes in bacterial targets, which prevent the efficient action of the antibiotic, and reduction of the effective intracellular concentration of the antibiotic, achieved by different mechanisms described in Fig. 12.1B.

A formerly susceptible organism can acquire resistance by mutation[17] or by incorporating foreign DNA.[18] Mutation is the major cause of resistance during infections in the absence of a donor of antibiotic-resistance genes.[19] Mutations involved in the development of resistance produce structural changes in the targets (for instance, quinolone resistance due to mutations in genes coding for bacterial topoisomerases[20]), in the enzymes that activate the preantibiotic (e.g., resistance to isoniazid due to mutations in the *Mycobacterium* catalase gene[21]), or in antibiotic transporters. Mutations in regulatory elements are also relevant for acquiring resistance. They act by changing the level of expression of antibiotic transporters (for instance, the porin OprD2 that transports imipenem inside *P. aeruginosa* is not expressed in imipenem-resistant mutants[22]) or by increasing the expression of antibiotic-detoxifying systems, such as chromosomally encoded antibiotic inactivating enzymes[23] and MDR efflux pumps.[24] Mutation is also important for the evolution of antibiotic-resistance genes acquired by HGT.

Resistance can be also achieved as a consequence of incorporating DNA from other bacteria. Occasionally, this DNA recombines with homologous genes of the new host rendering novel mosaic genes that make the host resistant to antibiotics (e.g., the formation of recombinant penicillin-binding proteins[25,26]). Alternatively, resistance results from the acquisition of an element that confers resistance on its own.[5,27] Given that bacterial pathogens were susceptible to antibiotics prior to their use as therapeutic agents for treating infections,[28] one intriguing question concerns the origin of antibiotic-resistance determinants.

Since natural antibiotics are produced by environmental microorganisms,[29] it was earlier proposed that antibiotic-producing microorganisms would be the most likely source of resistance determinants.[30] The rationality of this proposal derives from the need of producers to protect themselves from the activity of their own antimicrobials. In addition, antibiotic producers and resistant organisms can coexist in the same habitat, indicating that the inhibitory action of antibiotics may have an ecological value.[31] Nevertheless, in the few occasions in which the direct origin of some specific resistance genes has been tracked, they were derived from bacterial species that do not produce antibiotics. This is the case for Qnr determinants, which contribute to

Downloaded From Vet-eBooks.com

plasmid-acquired resistance to quinolones. It is worth mentioning that quinolones are synthetic antibiotics, so that it was proposed that resistance could be achieved only by mutations and that the existence of quinolone-resistance genes that could be acquired by human pathogens would be unlikely. Contrary to this hypothesis, the existence of plasmids carrying quinolone-resistance genes was described in 1998 in *Enterobacteriaceae*[32] and, as stated earlier, these plasmids are currently disseminated among *Enterobacteriaceae*. The search for *qnr* elements in the chromosomes of fully sequenced bacteria has shown that these determinants are mainly present in aquatic bacteria, which do not produce antibiotics.[33] A high conservation of sequence and chromosomal neighborhood indicates that the analyzed aquatic bacteria have not acquired those elements from other bacteria, but rather those microorganisms are the source of Qnr determinants. Indeed, the origin of the *qnrA1* gene, the most abundant in *Enterobacteriaceae* plasmids, is *Shewanella algae*.[34] Similarly, *Kluyvera ascorbata*, a nonproducer organism, is the most likely origin of the β-lactamases belonging to the CTX-M family.[35]

Resistance genes in producers are likely detoxification elements needed to avoid the activity of the antibiotic. However, their function in nonproducers is less apparent. They might serve to resist the activity of inhibitory compounds produced by competitors in complex microbial ecosystems. But some of these elements have not evolved to specifically counteract the activity of antibiotic producers. For instance, *Enterobacteriaceae* have harbored chromosomal β-lactamases[36] for several hundred millions of years. However, the natural habitat of these bacterial species is the gut, an ecosystem that does not contain β-lactam producers and, thus, did not contain β-lactam antibiotics until the use of these compounds for the treatment of infections. Similarly, quinolones are among the most frequent substrates of MDR efflux pumps[37] despite the synthetic origin of these antibiotics. A suitable hypothesis for explaining the origin of these elements is that they have been selected to play other roles than resistance,[38] but their natural substrates present structural similarities to antibiotics currently used for therapy; they can detoxify bacteria from these drugs even though this is not their original function.

For instance, all bacterial species harbor in their genomes genes coding for MDR efflux pumps. These genes are highly conserved (the same MDR elements are present in all the strains of a given bacterial species) and redundant (bacterial species usually harbor several different MDR efflux pumps), indicating that they are ancient elements relevant for bacterial physiology.[24] Given their genomic redundancy and their overlap in substrate usage, it is unlikely that the original function of these elements would be resistance to the drugs currently used in the clinical setting. This is the case of AcrAB-TolC, the major MDR efflux pump in *E. coli* and *Salmonella*. This element can extrude, in addition to antibiotics, bile salts,[39] which are toxic compounds present in the natural habitat of these species, the gut. Similarly, other MDR efflux pumps can have a primary role in the trafficking of bacterial signals,[40] the response to plant-produced signals,[41] the detoxification of intracellular toxic intermediary metabolites,[42] and the response to nonantibiotic bacterial inhibitors, such as heavy metals[43] or solvents.[44] The analysis of the effectors triggering the expression of MDR efflux pumps may help to understand their functional role in nature. This is the case of SmeDEF, the

Downloaded From Vet-eBooks.com

most important determinant of antibiotic resistance in *Stenotrophomonas maltophilia*,[45] whose expression is induced by plant flavonoids (and not by antibiotics), and its activity is needed for the colonization of plants roots by *S. maltophilia*.[46]

All this indicates that the universe of elements that can confer resistance to a heterologous host upon their transfer by HGT is larger than previously thought. As an example, the study of the resistome of the human gut microbiota[47] indicates the existence of a large number of elements that are not disseminated among human pathogens and that can confer resistance despite this ecosystem is not known to contain antibiotic producers.

Given the large number of resistance genes present in natural ecosystems, it is paradoxical that the number and variability of HGT acquired−resistance determinants currently present in human bacterial pathogens are relatively low. This might indicate that there are some restrictions for the transfer of a potential resistance element to a human pathogen. The first barrier would be ecological connectivity. The presence of resistance elements has been demonstrated in bacteria from the deep terrestrial subsurface[48] and in the deep Greenland ice core,[49] where human pathogens are not expected. The probability of transfer of these elements to human pathogens will be very low, but chances will increase for bacteria whose natural habitats are closer to those of human pathogens. One example of this type of "reactors" for resistance might be waste-water treatment plants, where human-linked microbiota (recipients of resistance genes) can get in contact with environmental microorganisms (potential donors) in the presence of residues of antibiotics that act as selecting agents.[8]

A second bottleneck for the transfer of a resistance gene will be its integration in an efficient dissemination vector or in a bacterial epidemic clone, which will allow a fast spread for the resistance determinant.[12,50]

The third bottleneck consists on the fitness costs associated with the acquisition of resistance.[51] It is generally accepted that the development of resistance by formerly susceptible bacteria might confer a metabolic burden such that resistant populations might be outcompeted by susceptible ones in the absence of antibiotic selective pressure. It is worth mentioning that, unless the fitness costs are unaffordable, their relevance during antibiotic treatment will be negligible because in these conditions being resistant is a prerequisite for sustaining an infection.[52] Nevertheless, fitness costs may be highly relevant for the persistence and spread of resistance in the absence of selection (for instance, in nonclinical, natural ecosystems). It has been described that the fitness costs associated to antibiotic resistance might be different depending on the mechanisms involved.[53] Furthermore, fitness costs can be compensated by secondary mutations in the bacterial genome[54] or by the rewiring of the bacterial metabolism in the resistant mutant.[55] The acquisition of a resistance determinant by a gene-recruitment element, such as an integron, which might harbor other resistance elements, might allow the maintenance of resistance by coselection events. Finally, the incorporation of a resistance determinant into a plasmid encoding toxin−antitoxin systems allows the persistence of resistance even in the absence of selective pressure.[56]

Overcoming all these obstacles is a necessary condition for the establishment of a resistance determinant in a bacterial population but it is not a sufficient one. Several resistance determinants besides those already disseminated among human pathogens

Downloaded From Vet-eBooks.com

might also overcome those bottlenecks. However, they are not disseminated currently. To explain this additional restriction, founder effects must be considered.[57] If one resistance element enters a formerly susceptible population through an efficient vector (or clone) and does not confer a high-fitness cost, it will soon spread under antibiotic selective pressure. Once this element is established, the antibiotic no longer exerts a selective pressure, because the bacteria are already resistant to it, so that the acquisition of a new resistance determinant will not represent an adaptive advantage. As a result, human antibiotic usage has rendered a strong increase in the prevalence of a few resistance elements that were previously present in the chromosomes of environmental microorganisms and are now located in gene-transfer units spreading not just in bacteria at clinical settings but also in environmental ecosystems.

The release of human pathogens harboring gene-transfer units containing resistance elements, eventually simultaneously with antibiotic-containing wastes, might have a deep impact on the evolution of the microbiota from natural ecosystems and this can also influence the evolution of clinically relevant mechanisms of antibiotic resistance.[8] As an example, the same antibiotic-resistance elements currently present in human pathogens can be found in wild animals[58] or in environmental locations without a history of antibiotic pollution.[59] Furthermore, the study of historical soils has demonstrated that the introduction of antibiotics has produced an increase in the prevalence of specific resistance determinants in environmental, nonclinical ecosystems.[60]

The recent use (in evolutionary terms) of antibiotics by humankind has produced a strong enrichment in the distribution of a few specific antibiotic-resistance elements in clinical and nonclinical ecosystems. The impact of this enrichment in specific genes, and eventually bacterial clones, on the composition and activity of the microbiosphere remains to be fully understood. Given that natural ecosystems are the source of resistance genes[61] and the reservoirs for their maintenance,[62] more studies on the ecological behavior of resistance in nonclinical habitats are required to unveil how these changes might impact the acquisition of antibiotic resistance by human pathogens.

3. Evolution of Antibiotic-Resistance Genes

3.1 Antibiotic-Resistance Genes as Targets of Evolution

The evolution and spread of antibiotic-resistance pathogens represent a unique opportunity for observing evolution in real time, and therefore a meeting point between researches from many different disciplines. The evolution of antibiotic resistance is a consequence of the selection of resistant organisms with particular genomic, physiological, or ecological abilities.

After the initial views that antibiotic-resistance genes had their origin in the environment,[30] likely in antibiotic producers, it was later accepted that genes encoding mechanisms of resistance or their precursors arose in potentially any bacteria, in most cases as house-keeping genes involved in the physiological functions required for daily bacterial life. Examples such as GadA and GadB proteins (glutamate

Downloaded From Vet-eBooks.com

decarboxylase) as well as AmpC and HdeB proteins, which increase ampicillin resistance in *E. coli*,[63] show the possibilities of different evolutionary pathways for developing antibiotic resistance in bacteria. Remarkably, proteins such as AmpC can confer resistance without further evolution, despite their presence in *Enterobacteriaceae* and that the gut is not known to harbor β-lactam producers, indicating that resistant phenotypes can occur and even evolve in the absence of antibiotic selection; conversely, antibiotics may influence the evolution of bacterial functions associated to the adaptation to particular environments. In any case, it is essential to understand that there is a wealth of potential mechanisms of resistance contained in bacterial chromosomes and in mobile genetic elements. In this section, we illustrate a number of issues related to the evolution of genes directly involved in antibiotic-resistance phenotypes.

The main mechanisms of gene variation leading to variation and diversification of antibiotic-resistance genes are mutation, recombination, and amplification. The frequency of these mechanisms is variable in normal populations, being typically from 10^{-9} to 10^{-6} in the case of mutation, from 10^{-7} to 10^{-13} for recombination, and from10^{-5} to 10^{-2} for tandem gene amplification.[64] However, the relative contribution of these factors depends on the bacterial species, the genomic background, and the treatment strategies.[65] For instance, mutation is the main mechanism for those microorganisms in genetic isolation (with few opportunities of recombination), such as rifampicin resistance in *M. tuberculosis*; on the contrary, recombination is more relevant for the resistance to β-lactams in *Streptococcus pneumoniae*. The known examples of amplification of antibiotic-resistance genes are overexpression of chromosomal AmpC in *Enterobacteriaceae* or insertion sequences upstream of antibiotic-resistance genes, such as ISEcp1 upstream of CTX-M-15.

Gene duplication also plays an important role in antibiotic resistance. An example is the duplication of *aphA1* involved in the tobramycin resistance during the therapy.[66] Moreover, the evolutionary results will be different, depending on exposed antibiotic concentrations. At low concentration of antibiotics (weak bottlenecks), many different low-level mechanisms of resistance could be selected (many times affecting different targets conferring low level of resistance), for instance, mutation in β-lactamase gene and porin-deficient subpopulations. However, at high concentrations (strong bottlenecks), only a few evolutionary pathways, or sometime only one mutant, are selected.[67]

Microorganisms acquire spontaneous mutations at relatively constant rate,[68] during the normal process of DNA replication. These mutations can randomly occur in any bacterial gene, including targets of antibiotics. When a single mutation in the target of antibiotic is sufficient to confer phenotypic resistance, the bacterial population carrying the mutated gene will be selected under antibiotic exposure. A good example is illustrated by fluoroquinolone resistance as the consequence of a single mutation in serine 83 of the A-subunit of the DNA gyrase, which provides a ciprofloxacin MIC of about 1 mg/mL in clinical isolates ripoll.[69] The most paradigmatic example in the arms race between antibiotic resistance in bacteria and the development of new compounds has been the evolution of TEM-1 β-lactamase, evolving by point mutation to different phenotypes, from narrow spectrum to ESBL (such as TEM-1 carrying G238G

Downloaded From Vet-eBooks.com

mutation) conferring resistance to oxyimino-cephalosporins or inhibitor resistance (such as TEM-1 carrying S130G mutation),[70] involved in β-lactam-plus β-lactamase inhibitor resistance.

A side effect of the use of antibiotics is the increase of mutation rate and consequently faster selection of resistant variants. There are two mechanisms to increase the mutation rate induced under antibiotic pressure: a loss of antimutator genes (or DNA repair genes), known as stable mutators, and the induction of SOS response and RpoS regulon (known as transient mutators), which increase the error-prone polymerases, such as DNA polymerase IV (dinB), and downregulation of MutS (caused by RpoS), a protein involved in the mismatch repair system (MMR).[71,72] Mutator strains are selected in fluctuating environmental, such as consecutive bottlenecks of different antibiotics in laboratory conditions,[73] and clinical settings. They have been described, for example, in clinical strains of *S. aureus*, *S. pneumoniae*, *H. influenzae*, *E. coli*, *P. aeruginosa*, or *S. maltophilia*.[74] Stable mutators present 10−1000-fold more chances of introducing changes in their DNA sequence during each replication cycle, and, therefore, under suboptimal growth rate they have an increased probability of selecting an advantageous in the cell survival, in shorter time than in nonmutator population (acquisition of antimicrobial-resistance) knapp2009.[60,61] From the evolutionary point of view, the mutator strains are selected by hitchhiking under recurrent selection pressures.

There are many studies about the role of hypermutation in the selection of resistance.[75] Mutators have been used to experimentally predict the emergence and selection of resistant variants;[69,76] however, they have a high biological cost in bacterial population replicating in stable environments.[77] On the contrary, transient mutators have lower fitness cost than stable mutators. Antimicrobial agents, such as fluoroquinolones or trimethoprim are DNA-damaging agents, inducing the SOS response. Among the cascade of overexpressed genes under SOS response are error-prone polymerases, such as DNA pol II (polB or dinA), pol IV (dinB), and pol IV (umDC). Moreover, these bactericidal antibiotics enhance the reactive oxygen species (ROS) in bacteria.[78] ROS is a potent inductor of damaged DNA, and consequently of SOS response.[79] In consequence, bactericidal antibiotics (bacteriostatic antibiotics are not inducers of hydroxyl radicals) have a double effect, inducing a process leading to cellular death by blocking their target (PBPs, DNA gyrase, and so on) while promoting the generation of genetic diversity and, therefore, of antibiotic resistance. An elegant model is fluoroquinolone resistance mediated by *qnr* genes: treatment with ciprofloxacin induces the SOS response, increasing the expression of error-prone polymerases (in a similar way to β-lactams or aminoglycosides), and promotes the cleavage of the LexA protein, a negative repressor of the qnrB2 gene, thus leading to QnrB overexpression. QnrB binds to DNA gyrase, protecting it from quinolone inhibition.[80]

Recombination is a powerful mechanism for the evolution of antibiotic-resistance genes. Antibiotic pressure can stimulate the HGT through SOS system. For instance, antibiotic-stimulated SOS induction can promote the transmission of *integrative conjugative elements* in *Vibrio cholerae* population.[81] Antibiotics can affect the intrachromosomal recombination. Intraorganismal gene recombination is also a powerful mechanism for the evolution of antibiotic-resistance genes, particularly

Downloaded From Vet-eBooks.com

relevant to the rapid spread of adaptive mutations within a genome when they occur in a copy of otherwise repeated homologous genes. This phenomenon, known as gene conversion,[82] increases, for instance, the efficiency of antibiotic-resistance mutations in *rrn* genes.

3.2 Adaptive Evolution of the Proteins Encoding for Antibiotic-Resistance Genes

The bifunctional enzyme AAC(6′)-Ib-cr gene is an aminoglycoside acetyltransferase evolved from AAC(6′)-Ib, with the capacity to hydrolyze aminoglycoside (the most prevalent gene that encodes aminoglycoside-modifying enzymes), in which the accumulation of two mutations, W102R and D179Y, extended the spectrum of activity to hydrolyzing fluoroquinolones.[83] The proposed model suggests that the D179Y mutation is responsible for the increase in the affinity of AAC(6′)-Ib-cr, whereas the W102R change could have an important role in the stabilization of the enzyme.[84] This antibiotic-resistance protein represents a remarkable example of enzymatic plasticity for acquiring new functions.[25]

The explanation for these evolutionary surprises is that many enzymes that have a main activity can also accept another subtract, as a consequence of their particular folding, around the active center. These proteins are identified as promiscuous enzymes and are widely distributed in all organisms, representing around 10% of the total enzymatic repertoire in bacteria,[85] but are more common in microorganisms frequently exposed to fluctuating environments, such as free-living organisms and pathogens. The secondary activities are generally multiple orders of magnitude lower than the native reactions, but they provide further potential starting points for novel functional adaptation. Enzyme promiscuity, therefore, provides a reservoir of candidates for evolutionary tinkering (resistome). The functional transition, based on the accumulation of mutations, from activity A (main) toward new activity B (previously promiscuous and residual), requires an overlapping of both activities through evolutionary intermediates.[86] A well-known example is β-lactamases (activity B), which evolved from PBPs (activity A), but even nowadays there are enzymes with both activities, such as PBP from *Mycobacterium smegmatis*.[87]

Depending on the time and intensity of selection, the final result of a promiscuous enzyme could be a bifunctional enzyme (activities A and B are equally efficient). This is the case of AAC(6′)/APH(2′).[88] In other cases, the overspecialization of activity B, implies the loss of activity A. This phenomenon is known as antagonistic pleiotropy,[76] and it has been widely described in β-lactamases, and is also described in other antibiotic families, such as tetracycline.[89]

The possibility of predicting the evolution of antibiotic-resistance evolution toward extended spectra of activity and to explore the capacities for acquiring new functions has been a field of interest for evolutionary biologists, microbiologists, and physicians. The most common experimental assay to predict the adaptability of antibiotic-resistance genes has been to expose a bacterial culture to increasing concentrations of the antibiotic.[69,90] This approach has several limitations. For instance, in general, only a single mutant is detected (the fittest in those particular experimental conditions),

Downloaded From Vet-eBooks.com

whereas in nature, experience has revealed that many trajectories may lead to resistance to a new antibiotic. Sometimes the most frequently selected mutant is not coincident with the mutant selected in the clinical setting.[91] These problems are indicating that this approach has a limited capacity to predict the evolution of antibiotic resistance.[92] Currently, the combination of crystallographic studies, bioinformatics, the deep sequencing, and ancestral reconstruction has allowed the identification of all intermediate evolutionary stages, to identify the selective forces driving the evolution and the different pathways and evolutionary constraints. All these advances have led to a more complete description of the evolutionary dynamics of antibiotic resistance.[65]

3.3 Defining Evolutionary Trajectories and Identifying Evolutionary Constriction Constraints As an Approach to Predict the Antibiotic Resistance: The Model of β-Lactamases

The adaptive potential of determinants conferring resistance to antibiotics has been described for most families of antibiotics, such as tetracyclines,[89] fluoroquinolones,[93] and aminoglycosides.[94] However, β-lactamases are the best model to understand the evolutionary potential of antibiotic-resistance elements,[95] because β-lactams are the most extensively used antibiotics in the clinical setting[96] and are the family for which largest number of chemical molecules have been developed. The simultaneous application of strong selective pressures and changing selectors (different β-lactams) has allowed the evolutionary radiation of β-lactamases.[70,97]

Resistance to β-lactam antibiotics can be attained by different mechanisms, such as the alteration of PBP's (for instance, in *S. pneumoniae*[98]), thus decreasing the affinity of β-lactam to the target. A second mechanism of resistance is the impermeability of the membrane,[99] decreasing the uptake of β-lactam into the bacterial cell. The most prevalent, widely distributed, and diverse mechanism of resistance to β-lactam is the hydrolysis of the β-lactam ring catalyzed by β-lactamases (www.lahey.org/studies/temtable.asp). The more than 1000 β-lactamases currently known are divided into four groups (A−D) according to their enzymatic properties and evolutionary relationships, class A being the most widely distributed.[95] In fact, the number of variant class A β-lactamases with clinical importance is enormous and they include about 498 OXA variants, 225 TEM, 195 SHV, 172 CTX-M, or 164 carbapenemases (distributed in different families, such as IMP, IMI, VIM, KPC, or NDM). Although the evolutionary root of these groups of β-lactamases originated in environmental bacteria, their subsequent evolution has likely occurred in clinical environments as the consequence of strong selective pressure by changing β-lactams. Nevertheless, according to random mutational experiments and deep sequencing, the number of distinct single-residue mutants for typical proteins is in the range of 103−104 and the number of all double mutants reaches the range of 106−108, much larger than the number of variants detected in nature.[100]

Although the diversity within TEM enzymes is high, affecting to 32% (92/286) of the amino acid positions, several authors have demonstrated that only 13−16% of

Downloaded From Vet-eBooks.com

positions in TEM-1 β-lactamase do not tolerate substitutions, being critical or drastically reducing the hydrolytic activity of this enzyme.[101,102] Several reasons may explain why only a limited number of mutants are possible.

TEM-1 β-lactamase was mutagenized in all the positions, but no change increasing the MIC to ampicillin was observed.[67] On the contrary, 2% of all changes in TEM-1 increased the activity on cefotaxime. One of them was R164 H/S/C, which simultaneously reduced the enzymatic stability.[103] Therefore, the selection of R164 H/S/C requires second-order mutations, known as global suppressors, such as M182T[102] or L201P,[104] which restore the thermostability.[105] This type of compensatory mutations have been described also in other β-lactamases, such as A77V in CTX-M[91,106] or R275L−N276D in SHV.[107] However, these mutations do not contribute to increase the β-lactam resistance when they are acquired in the first step.[108,109] Therefore, the phenotypic effect of these second-order mutations would depend on previous genetic background (epistasis).

Another example of evolutionary constrictions in TEM-1 affecting the R164 H/S/C is the negative reciprocal sign epistasis (exclusion effect) with other mutation involved in resistance to ceftazidime, such as G238G,[110] but each of them will facilitate different second-order mutations, suggesting different and incompatible evolutionary trajectories. Therefore, epistasis is increasingly recognized as a major constraint in evolution, which restricts accessible trajectories and can lead to different evolutionary outcomes.[86,108,109]

4. Limitations to Adaptation and the Cost of Resistance

4.1 The Genetics of Adaptation

Genetic variability in a population does not increase inevitably along time, since it is the result of factors acting in opposite directions: some processes introduce new variation in the populations while others remove it. Two main processes deplete variation from bacterial populations: selection and drift. By increasing the proportion of cells that carry particular, high-fitness variants, selection may transitorily reduce genetic variability in populations, while the effect of drift is continuous and equally affects all variants in the population, regardless of their effect on fitness.

Fitness can be defined as the relative capacity of bacteria to survive and reproduce within an infected individual and to spread to infect others. The epidemiological component of this definition emphasizes the need for considering all the levels at which fitness can be analyzed.[92] A very successful variant that can resist an antibiotic will be of very little relevance if it fails to be transmitted to other individuals. Both fitness components, intra- and interhost, are usually correlated, but this is not necessarily so. Evaluation of intrahost components can be approached using in vitro systems, but the transmission fitness is only possible from epidemiological observations.

Fitness is not a fixed property of individuals or groups: it is contextual and it can change when the environment or the genetic background are altered. This is readily

Downloaded From Vet-eBooks.com

exemplified by the cost of resistance, the reduction of the fitness of antibiotic-resistant bacteria in the absence of the drug. In the presence of the antibiotic, and occasionally in that of others as well, the increase in fitness associated to survival, reproduction, and transmission reveals the environmental dependence of the concept. Similarly, compensatory mutations can alter the fitness value of a certain resistance mutation by modifying the genetic context in which they are expressed. Naturally, both genetic and environmental changes can interact in synergistic or antagonistic ways, making more difficult the prediction of the phenotypic value under particular combinations of the two components.

Antibiotic-resistant mutants, as well as strains carrying resistance plasmids, can be fitter than wild-type strains in the presence of subinhibitory concentrations of antibiotics.[111,112] Upon these conditions, the selection of resistant mutants at very low concentrations of antibiotics, which comprises the range of concentrations between MIC and the mutant preventive concentration is feasible.[113,114] This situation opens the possibility for the selection of antibiotic-resistant mutants at places, such as waste-water treatment plants, where the concentration of the antibiotic is low but stays long enough to allow the displacement of the wild-type susceptible population by the resistant one. Similarly, a given population may be resistant to higher concentrations of an antibiotic or the necessary concentration of this to inhibit the bacterial growth completely (MIC value) can be higher. The values of these variables are often taken as indirect measures of fitness and they usually correlate with increased risk or potential harm in the clinical practice. A higher dose of an antibiotic may have serious side effects or may not be easily tolerated by some patients, thus posing at higher risk their survival from an infection.

Genetic drift is the result of the sampling process that occurs in every population in which the total number of individuals is limited. This limit can be very high (millions or billions, in the case of bacterial populations) or very low, as when an individual is infected by a single bacterial cell, as in some tuberculosis (TB) infections. In the former case, the reduction in genetic variability is almost imperceptible and it is easily compensated by the continuous generation of new genetic variation. On the contrary, extreme reductions in population size, especially during the transmission from one infected host to a new one, result in a drastic elimination of genetic variability after which only a few of the initially present variants are represented in the newly established population. In this case, the variants that originate the new population are drawn at random from those initially present, and the particular variants transmitted are not necessarily associated with increased fitness.

Although usually overlooked in the study of genetic variation in microorganisms, the neutral theory of molecular evolution[115,116] sustains that most variation at the molecular level does not have an impact on fitness and, as a consequence, is neutral in terms of natural selection. The original proposal was expanded[117] by incorporating the evolutionary consequences of slightly deleterious mutations whose fate does not depend exclusively on their relative fitness but also on the size of the population where they arise. Stochastic processes, usually associated with genetic drift, will dominate the fate of these mutations if effective population size is lower than the reciprocal of the corresponding selection coefficients. When population sizes or selection coefficients

Downloaded From Vet-eBooks.com

are larger and when the previously mentioned inequality no longer holds, then deterministic processes will dominate, and selection will be the main evolutionary force in the population. Given the large population sizes associated with bacteria, it is usually considered that genetic drift is not as important as selection in determining evolutionary change in bacterial populations. But this is not the case during transmission or during chronic infection. Effective sizes for pathogenic bacterial populations have been estimated to be much lower than for free-living bacteria.[118] This implies that stochastic factors may have an important role in the evolution of bacterial pathogens at this level. One additional, often overlooked, aspect of the quasineutral theory is that it also applies to slightly favorable mutations. While some mutations may confer increased fitness, their dynamics (stochastic or deterministic) will be determined by the relationship between the effective population size and the selection coefficient: a slightly advantageous mutation may easily disappear from a small population while it will likely increase in frequency in a large one.

The interplay between selection and drift can have consequences and leave imprints at different levels. The study of the evolution of CTX-M β-lactamases toward higher MIC values for cefotaxime and ceftazidime[108] demonstrates that some critical steps in some of the evolutionary trajectories revealed in the analysis were only possible if drift had played an important role, since the fitness of a necessary new genotype in a pathway was lower than that of the preceding variant. Apart from invoking evolution in alternative environments (with different fitness landscapes than those considered), this is only possible by the action of stochastic factors, among which drift is the major player. At a different level, reduced population sizes in *M. tuberculosis* may explain the higher relative rates of nonsynonymous substitutions in their genes when compared with other free-living bacteria.[119] In consequence, although selection may be the dominant factor in the evolution of bacterial populations, explaining almost perfectly the observed dynamics of antibiotic resistance in the presence of the selective drug, other evolutionary processes cannot be dismissed completely as irrelevant. Since these dynamics do not depend on fitness advantages, it is not necessary to invoke a cost of adaptation in every case and, most especially, in the absence of antibiotic.

4.2 From Genotype to Phenotype: The Many Ways Toward Fitness Compensation

While it is true that there are examples of drug-resistance mutations with no associated fitness cost, it is clear from their usually low frequencies that a fitness cost in the absence of the drug tends to be associated with resistance. This observation has led some researchers to argue that the removal of antibiotics will leave room to drug-susceptible strains and that these will outcompete those harboring drug-resistance mutations.[120] Although the strategy of drug removal seems to have some success in particular settings,[121,122] other factors apart from the total amounts of drug influence the frequency of drug resistance. One of these factors, as shown by several clinical, experimental, and epidemiological data, is compensation of fitness costs, so that the advantage of drug-susceptible strains in an antibiotic-free environment is reduced or even disappears.[123] As discussed previously, the fitness cost can be ameliorated

Downloaded From Vet-eBooks.com

through reversion, which is a very unlikely process.[9] It is more likely that, in the absence of the antibiotic, the low-fit drug-resistant strains either become extinct or find ways to recover fitness while keeping a drug-resistant phenotype. This process is usually known as compensation. Compensation is much more likely than reversion because there are usually many more loci that potentially can restore, at least partially, fitness costs. These loci can be in the same gene harboring the drug-resistance (intragenic) mutation, in other genes that somewhat interact with the drug resistance–mutated gene (intergenic), in plasmids, or in another chromosome, depending on the mechanisms giving more chances to compensate than to revert a drug-susceptible phenotype.[124]

Mechanisms leading to compensation of drug resistance can be grouped in three categories: (1) those based on chromosomal compensatory mutations, (2) those based on some kind of regulation alteration of the expression, and (3) those based on the so-called bypass mechanisms. Chromosomal mutations leading to compensation represent the case in which fitness loss is compensated by a second, or more, mutations. These mutations can occur either in the same protein affected by the drug-resistance gene (intragenic mutation) or in other proteins that interact with it (intergenic mutations). But the complexity of the compensation mutational pathways can go far beyond the accumulation of one or two mutations. Marcusson et al.[125] showed how in isogenic, lab-constructed strains of *E. coli* resistant to fluoroquinolones sometimes higher fitness effects are only attainable when four or five mutations are combined in the same strain and always depend on the loci mutated. Interaction between mutations can occur also among drug-resistance mutations for different antibiotics, sometimes leading to a higher (positive) or lower (negative) fitness than the mere sum of their individual effects, a phenomenon usually known as epistasis. Positive epistasis can explain why the frequency of high-cost drug-resistant strains in clinical settings is higher than expected.

A clear example of these epistatic interactions is shown by Trindade et al.[126] They introduced mutations to different drugs in isogenic strains thus creating MDR strains and focused on combinations of drug-resistant mutations to rifampicin (*rpoB* gene), nalidixic acid (*gyrA*), and streptomycin (*rpsL*). They found that several combinations of these mutations led to fitter-than-expected mutants. Furthermore, these mutations were not gene- but allele-specific and therefore epistasis and compensation depended on combinations of particular codon changes. In some cases, the double mutants not only were fitter-than-expected but also were fitter than at least one of the two individual mutants. This phenomenon is called sign epistasis and means that there is not only amelioration of the fitness cost between drug-resistant mutations (positive epistasis) but also compensation leading to partial restoration of fitness. It is interesting that combinations of *gyrA*, *rpsL*, and *rpoB* drug-resistance mutations have been shown to be present in different bacterial backgrounds, which suggests that epistasis among drug-resistance mutations can be present in many pathogens. In fact, multidrug clinical-resistant strains of *M. tuberculosis* have been reported to have higher fitness than their rifampicin-susceptible counterparts, thus indicating that compensation during treatment and/or epistatic effects between different drug-resistance mutations ameliorate, or even revert, the fitness cost of individual changes.[127]

Downloaded From Vet-eBooks.com

Another way to compensate for the fitness cost of drug-resistance mutations is at the level of gene expression.[124] There are examples of upregulation, through mutation, of a gene to counteract the negative effects on the expression of drug-resistance mutations, processes usually known as bypass mechanisms. The most typical example is that of *KatG* and the upregulation of *ahpC* in *M. tuberculosis*bib128.[128] Isoniazid is a prodrug and it needs the catalase−peroxidase activity of *katG* to become active. Mutations in *katG* confer resistance to isoniazid. A whole spectrum of mutations altering KatG function has been described,[129] and because the gene has an important role in the bacterial response to oxidative stress, it has been assumed that all of them have an associated fitness cost. It has been reported that the upregulation of the *ahpC* gene due to a mutation in its promoter can partially compensate for the loss of activity of *katG*, although there are some conflicting reports.[130]

Work reported in 2015 has shown that, occasionally, fitness costs are not detectable by current methods based on growth competitions, because resistant bacteria can activate alternative energy resources to cope with the costs associated to resistance.[5] However, even then the acquisition of resistance may produce relevant changes in bacterial physiology, including different traits involved in virulence.[131] A full understanding of the effects of resistance on bacterial fitness requires exploring more complex models than those based on growth in rich medium, currently the most popular ones.[92]

Another common way to increase the expression of a particular product is by gene duplication and amplification (GDA),[132] which may exhibit resistance to many antibiotics. However, it has been shown to be also a way of compensating for the fitness loss associated to resistance. GDA as a compensatory mechanism has been demonstrated more clearly in experimental evolution tests with *Salmonella enterica*.[133] Tandem duplications of the *metZ* and *metW* genes compensate for the loss of methionyl-tRNA formyl transferase by increasing levels of the nonformylated tRNA inhibitor, the one used by eukaryotes for translation initiation.

4.3 *Beyond Model Organisms: Epidemiological and Experimental Fitness Cost in* Mycobacterium tuberculosis

Experimental evolution with model microorganisms has been a successful approach to test evolutionary hypotheses. These experiments allow studying evolution in real-time producing accurate measures of key parameters, such as fitness, generation times, population sizes, or mutation rates.[134] Drug resistance can be approached within an evolutionary framework given that antibiotics are the main evolutionary pressure a microorganism can face jointly with the host's immune system.

A paradigmatic, or even extreme, case in this respect is *M. tuberculosis*, the causative agent of TB, with a colony-forming time of 3−4 weeks and which requires working in BSL3 facilities. This is why alternative model organisms, such as *M. smegmatis*, are frequently used to test hypotheses in TB research. Experimental work on drug resistance with *M. tuberculosis* has been successfully completed, corroborating many conclusions drawn from model organisms and justifying a constant feedback between model organism and real pathogens. A clear example is the

Downloaded From Vet-eBooks.com

evolution of drug resistance to rifampicin. Rifampicin targets the β-unit of the DNA-dependent RNA polymerase of microorganisms (encoded by the *rpoB* gene) by competing for the union to DNA and inhibiting RNA synthesis.[135] Therefore, it is a wide-spectrum antibiotic as it affects, with different efficiencies, many bacteria. Early work with *E. coli*[136] and other bacteria identified homologous positions mutated in drug-resistant strains, both in experimental and clinical settings, something expected given the high conservation of the *rpoB* gene among bacteria. A screening of gene mutations in *rpoB* from clinical strains of *M. tuberculosis* also identified many of them, as well as other mutations.[137] However, these mutations varied in frequency, suggesting a possible difference in the degree of resistance conferred and/or their associated fitness. Experimental evolution of two different lineages of *M. tuberculosis* revealed the existence of two main factors affecting drug-resistance fitness cost in this species: the genetic background of the strain and specific codon mutations.[138] Different codon mutations were found to have different fitness costs. Furthermore, these fitness costs varied between two lineages of *M. tuberculosis*, although in both cases the change S531L was the one associated with less fitness reduction. Mutations with lower fitness costs were found to be the most frequent among clinical strains, suggesting a correlation between "in vitro" and epidemiological fitness cost. Finally, the fitness of paired isolates of RIF^s and RIF^r strains from 10 different patients who converted to drug resistance during treatment were screened showing not only comparable results to the experimental findings but also cases in which the fitness RIF^r strain was higher than that of the RIF^s counterpart. Whole-genome sequencing of those isolates in parallel to the serial passage of rifampicin-resistance isolates in the absence of antibiotic have allowed to identify rifampicin resistance compensatory mutations in two subunits of the polymerase, *rpoA* and *rpoC*.[139] Furthermore, those mutations were common among the highly successful MDR clones of high-burden multidrug-resistance countries suggesting that they allow those strains with the mutations to be better transmitted.[139] Later studies have corroborated that mutations in *rpoA* and *rpoC* are associated to MDR strains involved in large outbreaks,[140,141] confirming their role in the successful transmission of MDR strains at the population level.[142,143]

5. Can the Evolution of Antibiotic Resistance be Predicted?

Conventional scientific wisdom dictates that evolution is a process that is sensitive to many unforeseeable events and influences and, therefore, is essentially unpredictable. On the other hand, considering the tremendous amounts of knowledge gained about bacterial genetics and genomics; population genetics and ecology of bacterial organisms; and their subcellular elements involved in HGT, we should consider the possibility of predicting the evolution of bacterial populations and traits[50] similar to weather forecasts, with higher probabilities of success in the closer and more local frames. Indeed, there is a *local* evolutionary biology based on local selective constraints that shape the possible local trajectories, even though in our global world some of these locally originated trends might result in global influences. In the case of

Downloaded From Vet-eBooks.com

adaptive functions (as antibiotic-resistance genes in pathogenic bacteria), some of the elements whose knowledge is critical for predicting evolutionary trajectories are: (1) the origin and function of these genes in the source environmental bacterial organisms; (2) their ability to be captured (mobilized) by different genetic platforms and to integrate in particular mobile genetic elements; (3) the ability of these mobile genetic elements to be selected, transferred, and spread among bacterial populations; (4) the probability of intrahost mutational variation and recombination; (5) the probability of recombination events among these and other mobile elements, with consequences in selectable properties and bacterial host-ranges, (6) the original and resulting fitness of the bacterial clones in which the new functions are hosted, including their colonization power and capacity to spread in an epidemic form; (7) the results of the interactions between these bacterial hosts and the microbial environments in which they are inserted; and (8) the selective events, such as the patterns of local antibiotic consumption or industrial pollution, and, in general, the structure of the environment that might influence the success of particular genetic configurations in which the adaptive genes are hosted. Dealing *simultaneously* with all these sources of evolutionary variation is certainly a challenge at present.

Such a type of complex structure has evolved along all hierarchical levels of biology, creating specific "Chinese-boxes" or "Russian-dolls" patterns of stable (preferential) combinations; for instance, encompassing bacterial species, phylogenetic subspecific groups, clones, plasmids, transposons, insertion sequences, and genes encoding adaptive traits. Assuming a relatively high frequency of combinatorial events, the existing trans-hierarchical combinations are probably the result of the local availability of the different elements (pieces) in particular locations (local biology), the local advantage provided by particular combinations, and also the biological cost in fitness of some of them. More research is needed to draw the interactive pattern of biological pieces in particular environments (grammar of affinities). Such a complex framework required for predicting evolutionary trajectories will be analyzed (and integrated) by considering heuristic techniques for the understanding of multilevel selection. The application of new methods, based on covariance, and contextual analysis, for instance using Price's equation,[144] should open an entirely new synthetic way of approaching the complexity of the living world.

6. Conclusions and Perspectives

In the absence of new antibiotics, most efforts have focused in protecting the few current ones that maintain activity, trying to reduce their strong selective effects by reducing antibiotic consumption in animals and humans while maintaining their efficiency. In a number of countries, this collective policy has proven insufficient. It has been proposed that the control of antibiotic exposure should be considered by society as an individual-based attitude to reduce individual risks, using similar approaches to those for controlling tobacco-associated diseases, hypercholesterolemia, or hypertension.[145] Reductions in the host-to-host transmission of resistant organisms through innovative approaches trying to influence the ecology and evolution of

Downloaded From Vet-eBooks.com

resistant organisms might represent alternative ways to limit the spread of antibiotic resistance in the microbiosphere. In this respect, the possibility of applying in the future eco-evo drugs—drugs acting *not* to cure the individual patient but to "cure" specific environments from antibiotic resistance, and to prevent or weaken the evolutionary possibilities (the evolvability) of the biological elements involved in it— should be considered. In other words, this strategy proposes to combat (decontaminate, deevolve) resistance not in infected patients, but rather in the whole population, including infected and noninfected people alike, as it occurs in hospitals, nurseries, elderly facilities, and so on. By extension, other environments that can be successfully treated are farms, fish factories, or sewage facilities. Indeed, the notion of "ill environment" should be increasingly encouraged, and medical-like approaches might be increasingly applied to prevent and cure biologically altered environments.[6]

The targets of these future drugs, some of them in early development, are not only resistant, "high-risk" clones but also the interbacterial transmissibility, the maintenance of bacterial plasmids, and integrative–conjugative elements carrying resistance, the ability of transposons and integrons to move between genomes, or the mechanisms of bacterial adaptation to antibiotic stress, including control of mutation and recombination rates.

Glossary

Founder effect The random change in genetic composition of a population due to an extreme reduction in its size during a colonization or infection episode.

Genetic drift The random change in the genetic composition of a population due to its finite size. Every population experiences genetic drift but its effects, a reduction in genetic variation eventually leading to fixation of a variant, are more intense, both in magnitude and speed, the smaller its population size.

Mutator strains Bacterial strains with an increased mutation rate usually due to a defective mismatch repair system.

Pleiotropic antagonism The effect of a gene on two different traits with opposite consequences on fitness.

Resistome The set of antibiotic-resistance genes or proteins found in a given environment.

List of Abbreviations

ESBL	Extended-spectrum beta-lactamases
GDA	Gene duplication and amplification
HGT	Horizontal gene transfer
MDR	Multidrug resistance
MGE	Mobile genetic element
MIC	Minimal inhibitory concentration

Downloaded From Vet-eBooks.com

MRSA	Methicillin-resistant *Staphylococcus aureus*
R_0	Basic reproductive number
RIFr	Rifampicin resistant
RIFs	Rifampicin susceptible
TB	Tuberculosis

Acknowledgments

We would like to thank Michel Tibayrenc for providing the opportunity to contribute to this volume and to the following financial supporters of our research: Instituto de Salud Carlos III REIPI RD12/0015 (J.L.M.) FIS PI1200567 (J.C.G.), Madrid Autonomous Community S2010/BMD2414 (J.L.M.), MINECO BIO2014-54507-R and JPI Water StARE JPIW2013-089-C02-01 (J.L.M.), BFU2014-58566R (F.G.C.), EU HEALTH-F3-2011-282004 (J.L.M.), Ramón y Cajal Spanish research grant RYC-2012-10627, MINECO research grant SAF2013-43521-R, and the European Research Council (ERC) (638553-TB-ACCELERATE) to IC.

References

1. Vatopoulos A. High rates of metallo-beta-lactamase-producing *Klebsiella pneumoniae* in Greece−a review of the current evidence. *Euro Surveill* 2008;**13**:8023.
2. Wright A, Zignol M, Van DA, Falzon D, Gerdes SR, Feldman K, et al. Epidemiology of antituberculosis drug resistance 2002−07: an updated analysis of the global project on anti-tuberculosis drug resistance surveillance. *Lancet* 2009;**373**:1861−73.
3. de Lencastre H, Tomasz A. In: Baquero F, Nombela C, Cassell GH, Gutiérrez-Fuentes JA, editors. *Multiple stages in the evolution of methicillin-resistant* Staphylococcus aureus. American Society for Microbiology ASM; 2008. p. 333−46.
4. Souli M, Galani I, Giamarellou H. Emergence of extensively drug-resistant and pandrug-resistant Gram-negative bacilli in Europe. *Euro Surveill* 2008;**13**:19045.
5. Martinez JL, Coque TM, Baquero F. What is a resistance gene? Ranking risk in resistomes. *Nat Rev Micro* 2015;**13**:116−23.
6. Baquero F. Predictions: evolutionary trajectories and planet medicine. *Microb Biotech* 2009;**2**:130−2.
7. Martinez JL. The role of natural environments in the evolution of resistance traits in pathogenic bacteria. *Proc R Soc B* 2009;**276**:2521−30.
8. Baquero F, Martinez JL, Canton R. Antibiotics and antibiotic resistance in water environments. *Curr Opin Biotechnol* 2008;**19**:260−5.
9. Levin BR, Perrot V, Walker N. Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. *Genetics* 2000;**154**:985−97.
10. Bergstrom CT, Lo M, Lipsitch M. Ecological theory suggests that antimicrobial cycling will not reduce antimicrobial resistance in hospitals. *Proc Natl Acad Sci USA* 2004;**101**:13285−90.
11. Sundqvist M, Geli P, Andersson DI, Sjölund-Karlsson M, Runehagen A, Cars H, et al. Little evidence for reversibility of trimethoprim resistance after a drastic reduction in trimethoprim use. *J Antimicrob Chemother* 2010;**65**:350−60.
12. Baquero F. From pieces to patterns: evolutionary engineering in bacterial pathogens. *Nat Rev Micro* 2004;**2**:510−8.

Downloaded From Vet-eBooks.com

13. Tamae C, Liu A, Kim K, Sitz D, Hong J, Becket E, et al. Determination of antibiotic hypersensitivity among 4,000 single-gene-knockout mutants of *Escherichia coli*. *J Bacteriol* 2008;**190**:5981−8.
14. Alvarez-Ortega C, Wiegand I, Olivares J, Hancock RE, Martínez JL. The intrinsic resistome of *Pseudomonas aeruginosa* to β-lactams. *Virulence* 2011;**2**:144−6.
15. Dötsch A, Becker T, Pommerenke C, Magnowska Z, Jänsch L, Häussler S. Genomewide identification of genetic determinants of antimicrobial drug resistance in *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* 2009;**53**:2522−31.
16. Fernández L, Álvarez-Ortega C, Wiegand I, Olivares J, Kocíncová D, Lam JS, et al. Characterization of the polymyxin B resistome of *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* 2013;**57**:110−9.
17. Martínez JL, Baquero F. Mutation frequencies and antibiotic resistance. *Antimicrob Agents Chemother* 2000;**44**:1771−7.
18. Davies J. Inactivation of antibiotics and the dissemination of resistance genes. *Science* 1994;**264**:375−82.
19. Maciá MD, Blanquer D, Togores B, Sauleda J, Pérez JL, Oliver A. Hypermutation is a key factor in development of multiple-antimicrobial resistance in *Pseudomonas aeruginosa* strains causing chronic lung infections. *Antimicrob Agents Chemother* 2005;**49**:3382−6.
20. Piddock LJV. Mechanisms of fluoroquinolone resistance: an update 1994−1998. *Drugs* 1999;**58**:11−8.
21. Zhang Y, Heym B, Allen B, Young D, Cole S. The catalase-peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*. *Nature* 1992;**358**:591−3.
22. Yoneyama H, Nakae T. Mechanism of efficient elimination of protein D2 in outer membrane of imipenem-resistant *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* 1993;**37**:2385−90.
23. Nelson EC, Elisha BG. Molecular basis of AmpC hyperproduction in clinical isolates of *Escherichia coli*. *Antimicrob Agents Chemother* 1999;**43**:957−9.
24. Martínez JL, Sánchez MB, Martínez-Solano L, Hernández A, Garmendia L, Fajardo A, et al. Functional role of bacterial multidrug efflux pumps in microbial natural ecosystems. *FEMS Microbiol Rev* 2009;**33**:430−49.
25. Sibold C, Henrichsen J, König A, Martin C, Chalkley L, Hakenbeck R. Mosaic *pbpX* genes of major clones of penicillin-resistant *Streptococcus pneumoniae* have evolved from *pbpX* genes of a penicillin-sensitive *Streptococcus oralis*. *Mol Microbiol* 1994;**12**:1013−23.
26. Spratt BG, Bowler LD, Zhang QY, Zhou J, Smith JM. Role of interspecies transfer of chromosomal genes in the evolution of penicillin resistance in pathogenic and commensal *Neisseria* species. *J Mol Evol* 1992;**34**:115−25.
27. Davies JE. Origins, acquisition and dissemination. *Antibiotic Resist* 2008;**787**:15−27.
28. Datta N, Hughes VM. Plasmids of the same Inc. groups in Enterobacteria before and after the medical use of antibiotics. *Nature* 1983;**306**:616−7.
29. Waksman SA, Woodruff HB. The soil as a source of microorganisms antagonistic to disease-producing bacteria. *J Bacteriol* 1940;**40**:581.
30. Benveniste R, Davies J. Aminoglycoside antibiotic-inactivating enzymes in actinomycetes similar to those present in clinical isolates of antibiotic-resistant bacteria. *Proc Natl Acad Sci USA* 1973;**70**:2276−80.
31. Laskaris P, Tolba S, Calvo-Bado L, Wellington L. Coevolution of antibiotic production and counter-resistance in soil bacteria. *Environ Microbiol* 2010;**12**:783−96.
32. Martínez-Martínez L, Pascual A, Jacoby GA. Quinolone resistance from a transferable plasmid. *Lancet* 1998;**351**:797−9.

Downloaded From Vet-eBooks.com

33. Sánchez MB, Hernández A, Rodríguez-Martínez JM, Martínez-Martínez L, Martínez JL. Predictive analysis of transmissible quinolone resistance indicates *Stenotrophomonas maltophilia* as a potential source of a novel family of Qnr determinants. *BMC Microbiol* 2008;**8**:1.
34. Poirel L, Rodriguez-Martinez JM, Mammeri H, Liard A, Nordmann P. Origin of plasmid-mediated quinolone resistance determinant QnrA. *Antimicrob Agents Chemother* 2005;**49**: 3523−5.
35. Humeniuk C, Arlet G, Gautier V, Grimont P, Labia R, Philippon A. Beta-lactamases of *Kluyvera ascorbata*, probable progenitors of some plasmid-encoded CTX-M types. *Antimicrob Agents Chemother* 2002;**46**:3045−9.
36. Lindberg F, Normark S. Contribution of chromosomal β-lactamases to β-lactam resistance in enterobacteria. *Rev Infect Dis* 1986;**8**:S292−304.
37. Alonso A, Rojo F, Martinez JL. Environmental and clinical isolates of *Pseudomonas aeruginosa* show pathogenic and biodegradative properties irrespective of their origin. *Environ Microbiol* 1999;**1**:421−30.
38. Fajardo A, Martínez-Martín N, Mercadillo M, Galán JC, Ghysels B, Matthijs S, et al. The neglected intrinsic resistome of bacterial pathogens. *PLoS One* 2008;**3**:e1619.
39. Thanassi DG, Cheng LW, Nikaido H. Active efflux of bile salts by *Escherichia coli*. *J Bacteriol* 1997;**179**:2512−8.
40. Köhler T, van Delden C, Curty LK, Hamzehpour MM, Pechere JC. Overexpression of the MexEF-OprN multidrug efflux system affects cell-to-cell signaling in *Pseudomonas aeruginosa*. *J Bacteriol* 2001;**183**:5213−22.
41. Valecillos AM, Rodríguez Palenzuela P, López-Solanilla E. The role of several multidrug resistance systems in *Erwinia chrysanthemi* pathogenesis. *Mol Plant-Microbe Interact* 2006;**19**:607−13.
42. Aendekerk S, Diggle SP, Song Z, Høiby N, Cornelis P, Williams P, et al. The MexGHI-OpmD multidrug efflux pump controls growth, antibiotic susceptibility and virulence in *Pseudomonas aeruginosa* via 4-quinolone-dependent cell-to-cell communication. *Microbiology* 2005;**151**:1113−25.
43. Nies DH. Efflux-mediated heavy metal resistance in prokaryotes. *FEMS Microbiol Rev* 2003;**27**:313−39.
44. Ramos A, Hu DJ, Nguyen L, Phan KO, Vanichseni S, Promadej N, et al. Intersubtype Human Immunodeficiency Virus type 1 superinfection following seroconversion to primary infection in two injection drug users. *J Virol* 2002;**76**:7444−52.
45. Alonso A, Martinez JL. Expression of multidrug efflux pump SmeDEF by clinical isolates of Stenotrophomonas maltophilia. *Antimicrob Agents Chemother* 2001;**45**:1879−81.
46. García-León G, Hernández A, Hernando-Amado S, Alavi P, Berg G, Martínez JL. A function of SmeDEF, the major quinolone resistance determinant of *Stenotrophomonas maltophilia*, is the colonization of plant roots. *Appl Environ Microbiol* 2014;**80**:4559−65.
47. Sommer MOA, Dantas G, Church GM. Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* 2009;**325**:1128−31.
48. Brown MG, Balkwill DL. Antibiotic resistance in bacteria isolated from the deep terrestrial subsurface. *Microb Ecol* 2009;**57**:484−93.
49. Miteva VI, Sheridan PP, Brenchley JE. Phylogenetic and physiological diversity of microorganisms isolated from a deep Greenland glacier ice core. *Appl Environ Microbiol* 2004;**70**:202−13.
50. Martinez JL, Baquero F, Andersson DI. Predicting antibiotic resistance. *Nat Rev Micro* 2007;**5**:958−65.

Downloaded From Vet-eBooks.com

51. Andersson DI. The biological cost of mutational antibiotic resistance: any practical conclusions? *Curr Opin Microbiol* 2006;**9**:461–5.
52. Martinez JL, Baquero F. Interactions among strategies associated with bacterial infection: pathogenicity, epidemicity, and antibiotic resistance. *Clin Microbiol Rev* 2002;**15**: 647–79.
53. Balsalobre L, de la Campa AG. Fitness of *Streptococcus pneumoniae* fluoroquinolone-resistant strains with topoisomerase IV recombinant genes. *Antimicrob Agents Chemother* 2008;**52**:822–30.
54. Paulander W, Maisnier-Patin S, Andersson DI. Multiple mechanisms to ameliorate the fitness burden of mupirocin resistance in *Salmonella typhimurium*. *Mol Microbiol* 2007; **64**:1038–48.
55. Olivares J, Álvarez-Ortega C, Martínez JL. Metabolic compensation of fitness costs associated with overexpression of the multidrug efflux pump MexEF-OprN in *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* 2014;**58**:3904–13.
56. Moritz EM, Hergenrother PJ. Toxin-antitoxin systems are ubiquitous and plasmid-encoded in vancomycin-resistant enterococci. *Proc Natl Acad Sci USA* 2007;**104**: 311–6.
57. Martinez JL, Fajardo A, Garmendia L, Hernandez A, Linares JF, Martinez-Solano L, et al. A global view of antibiotic resistance. *FEMS Microbiol Rev* 2009;**33**:44–65.
58. Livermore DM, Warner M, Hall L, Enne VI, Projan SJ, Dunman PM, et al. Antibiotic resistance in bacteria from magpies *Pica pica* and rabbits *Oryctolagus cuniculus* from west Wales. *Environ Microbiol* 2001;**3**:658–61.
59. Pei R, Kim SC, Carlson KH, Pruden A. Effect of river landscape on the sediment concentrations of antibiotics and corresponding antibiotic resistance genes ARG. *Water Res* 2006;**40**:2427–35.
60. Knapp CW, Dolfing J, Ehlert PA, Graham DW. Evidence of increasing antibiotic resistance gene abundances in archived soils since 1940. *Environl Sci Technol* 2009;**44**: 580–7.
61. Martínez JL. Antibiotics and antibiotic resistance genes in natural environments. *Science* 2008;**321**:365–7.
62. Simões RR, Poirel L, Da Costa PM, Nordmann P. Seagulls and beaches as reservoirs for multidrug-resistant *Escherichia coli*. *Emerg Infect Dis* 2010;**16**:110–2.
63. Adam M, Murali B, Glenn NO, Potter SS. Epigenetic inheritance based evolution of antibiotic resistance in bacteria. *BMC Evol Biol* 2008;**8**:52.
64. Andersson DI, Hughes D. Microbiological effects of sublethal levels of antibiotics. *Nat Rev Micro* 2014;**12**:465–78.
65. Palmer AC, Kishony R. Understanding, predicting and manipulating the genotypic evolution of antibiotic resistance. *Nat Rev Genet* 2013;**14**:243–8.
66. McGann P, Courvalin P, Snesrud E, Clifford RJ, Yoon EJ, Onmus-Leone F, et al. Amplification of aminoglycoside resistance gene *aphA1* in *Acinetobacter baumannii* results in tobramycin therapy failure. *mBio* 2014;**5**:e00915–14.
67. Stiffler MA, Hekstra DR, Ranganathan R. Evolvability as a function of purifying selection in TEM-1 b-lactamase. *Cell* 2015;**160**:882–92.
68. Drake JW. A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci USA* 1991;**88**:7160–4.
69. Ripoll A, Baquero F, Novais Â, Rodríguez-Domínguez MJ, Turrientes MC, Cantón R, et al. In vitro selection of β-lactam plus β-lactamase inhibitor resistant variants in CTX-M β-lactamases: predicting the *in-vivo* scenario? *Antimicrob Agents Chemother* 2011;**55**: 4530–6.

Downloaded From Vet-eBooks.com

70. Guthrie VB, Allen J, Camps M, Karchin R. Network models of TEM β-lactamase mutations coevolving under antibiotic selection show modular structure and anticipate evolutionary trajectories. *PLoS Comput Biol* 2011;**7**:e1002184.
71. Gutierrez A, Laureti L, Crussard S, Abida H, Rodriguez-Rojas A, Blázquez J, et al. β-lactam antibiotics promote bacterial mutagenesis via an RpoS-mediated reduction in replication fidelity. *Nat Comm* 2013;**4**:1610.
72. Qin TT, Kang HQ, Ma P, Li PP, Huang LY, Gu B. SOS response and its regulation on the fluoroquinolone resistance. *Ann Trans Med* 2015;**3**.
73. Mao EF, Lane L, Lee J, Miller JH. Proliferation of mutators in a cell population. *J Bacteriol* 1997;**179**:417−22.
74. Marshall CG, Lessard IAD, Park IS, Wright GD. Glycopeptide antibiotic resistance genes in glycopeptide-producing organisms. *Antimicrob Agents Chemother* 1998;**42**:2215−20.
75. Chopra I, O'Neill AJ, Miller K. The role of mutators in the emergence of antibiotic-resistant bacteria. *Drug Res Updat* 2003;**6**:137−45.
76. Galán JC, Morosini MI, Baquero MR, Reig M, Baquero F. *Haemophilus influenzae* blaROB-1 mutations in hypermutagenic DampC *Escherichia coli* conferring resistance to cefotaxime and β-lactamase inhibitors and increased susceptibility to cefaclor. *Antimicrob Agents Chemother* 2003;**47**:2551−7.
77. Turrientes MC, Baquero F, Levin BR, Martínez JL, Ripoll A, González-Alba JM, et al. Normal mutation rate variants arise in a Mutator Mut S *Escherichia coli* population. *PLoS One* 2013;**8**:e72963.
78. Pena-Miller R, Laehnemann D, Jansen G, Fuentes-Hernandez A, Rosenstiel P, Schulenburg H, et al. When the most potent combination of antibiotics selects for the greatest bacterial load: the smile-frown transition. *PLoS Biol* 2013;**11**:e1001540.
79. Kohanski MA, DePristo MA, Collins JJ. Sublethal antibiotic treatment leads to multidrug resistance via radical-induced mutagenesis. *Mol Cell* 2010;**37**:311−20.
80. Da Re S, Garnier F, Guérin E, Campoy S, Denis F, Ploy M-C. The SOS response promotes *qnrB* quinolone-resistance determinant expression. *EMBO Rep* 2009;**10**:929−33.
81. Beaber JW, Hochhut B, Waldor MK. SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature* 2004;**427**:72−4.
82. Prammananan T, Sander P, Springer B, Böttger EC. RecA-mediated gene conversion and aminoglycoside resistance in strains heterozygous for rRNA. *Antimicrob Agents Chemother* 1999;**43**:447−53.
83. Maurice F, Broutin I, Podglajen I, Benas P, Collatz E, Dardel F. Enzyme structural plasticity and the emergence of broad-spectrum antibiotic resistance. *EMBO Rep* 2008;**9**:344−9.
84. Vetting MW, Park CH, Hegde SS, Jacoby GA, Hooper DC, Blanchard JS. Mechanistic and structural analysis of aminoglycoside N-acetyltransferase AAC 6′.-Ib and its bifunctional, fluoroquinolone-active AAC 6′.-Ib-cr variant. *Biochemistry* 2008;**47**:9825−35.
85. Martínez-Núñez MA, Poot-Hernandez AC, Rodríguez-Vázquez K, Perez-Rueda E. Increments and duplication events of enzymes and transcription factors influence metabolic and regulatory diversity in prokaryotes. *PLoS One* 2013;**8**:e69707.
86. Kaltenbach M, Tokuriki N. Dynamics and constraints of enzyme evolution. *J Exp Zool B Mol Dev Evol* 2014;**322**:468−87.
87. Bansal A, Kar D, Murugan RA, Mallick S, Dutta M, Pandey SD, et al. A putative low-molecular-mass penicillin-binding protein PBP of *Mycobacterium smegmatis* exhibits prominent physiological characteristics of DD-carboxypeptidase and beta-lactamase. *Microbiology* 2015;**161**:1081−91.

Downloaded From Vet-eBooks.com

88. Zhang W, Fisher JF, Mobashery S. The bifunctional enzymes of antibiotic resistance. *Curr Op Microbiol* 2009;**12**:505−11.
89. Linkevicius M, Sandegren L, Andersson DI. Potential of tetracycline resistance proteins to evolve tigecycline resistance. *Antimicrob Agents Chemother* 2016;**60**:789−96.
90. Drago L, De Vecchi E, Nicola L, Tocalli L, Gismondo MR. In vitro selection of resistance in *Pseudomonas aeruginosa* and *Acinetobacter* spp. by levofloxacin and ciprofloxacin alone and in combination with b-lactams and amikacin. *J Antimicrob Chemother* 2005;**56**: 353−9.
91. Novais A, Canton R, Coque TM, Moya A, Baquero F, Galan JC. Mutational events in ESBL-cefotaximases of the CTX-M-1 cluster involved in ceftazidime resistance. *Antimicrob Agents Chemother* 2008;**52**:2377−82.
92. Martínez JL, Baquero F, Andersson DI. Beyond serial passages: new methods for predicting the emergence of resistance to novel antibiotics. *Curr Op Pharmacol* 2011;**11**: 439−45.
93. Zhang G, Wang C, Sui Z, Feng J. Insights into the evolutionary trajectories of fluoroquinolone resistance in *Streptococcus pneumoniae*. *J Antimicrob Chemother* 2015;**70**: 2499−506.
94. Kramer JR, Matsumura I. Directed evolution of aminoglycoside phosphotransferase 3′. type IIIa variants that inactivate amikacin but impose significant fitness costs. *PLoS One* 2013;**8**:e76687.
95. Bush K. The ABCD's of β-lactamase nomenclature. *J Infect Chemother* 2013;**19**:549−59.
96. Goossens H. Antibiotic consumption and link to resistance. *Clin Microbiol Infect* 2009;**15**: 12−5.
97. Bush K. Alarming β-lactamase-mediated resistance in multidrug-resistant *Enterobacteriaceae*. *Curr Op Microbiol* 2010;**13**:558−64.
98. Tait-Kamradt AG, Cronan M, Dougherty TJ. Comparative genome analysis of high-level penicillin resistance in *Streptococcus pneumoniae*. *Microb Drug Resist* 2009;**15**:69−75.
99. Tran QT, Mahendran KR, Hajjar E, Ceccarelli M, Davin-Regli A, Winterhalter M, et al. Implication of porins in β-lactam resistance of *Providencia stuartii*. *J Biol Chem* 2010;**285**: 32273−81.
100. Figliuzzi M, Jacquier H, Schug A, Tenaillon O, Weigt M. Coevolutionary landscape inference and the context-dependence of mutations in beta-lactamase TEM-1. *Mol Biol Evol* 2016;**33**:268−80.
101. Huang W, Petrosino J, Hirsch M, Shenkin PS, Palzkill T. Amino acid sequence determinants of β-lactamase structure and activity. *J Mol Biol* 1996;**258**:688−703.
102. Jacquier H, Birgy A, Le Nagard H, Mechulam Y, Schmitt E, Glodt J, et al. Capturing the mutational landscape of the beta-lactamase TEM-1. *Proc Natl Acad Sci USA* 2013;**110**: 13067−72.
103. Petrosino JF, Palzkill T. Systematic mutagenesis of the active site omega loop of TEM-1 beta-lactamase. *J Bacteriol* 1996;**178**:1821−8.
104. Marciano DC, Pennington JM, Wang X, Wang J, Chen Y, Thomas VL, et al. Genetic and structural characterization of an L201P global suppressor substitution in TEM-1 β-lactamase. *J Mol Biol* 2008;**384**:151−64.
105. Brown NG, Pennington JM, Huang W, Ayvaz T, Palzkill T. Multiple global suppressors of protein stability defects facilitate the evolution of extended-spectrum TEM β-lactamases. *J Mol Biol* 2010;**404**:832−46.
106. Patel MP, Fryszczyn BG, Palzkill T. Characterization of the global stabilizing substitution A77V and its role in the evolution of CTX-M β-lactamases. *Antimicrob Agents Chemother* 2015;**59**:6741−8.

Downloaded From Vet-eBooks.com

107. Winkler ML, Bonomo RA. SHV-129: a gateway to global suppressors in the SHV β-lactamase family? *Mol Biol Evol* 2016;**33**:429−41.
108. Novais Â, Comas I, Baquero F, Cantón R, Coque TM, Moya A, et al. Evolutionary trajectories of beta-lactamase CTX-M-1 cluster enzymes: predicting antibiotic resistance. *PLoS Pathog* 2010;**6**:e1000735.
109. Weinreich DM, Delaney NF, DePristo MA, Hartl DL. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 2006;**312**:111−4.
110. Dellus-Gur E, Elias M, Caselli E, Prati F, Salverda ML, de Visser JAG, et al. Negative epistasis and evolvability in TEM-1 β-lactamase-the thin line between an enzyme's conformational freedom and disorder. *J Mol Biol* 2015;**427**:2396−409.
111. Andersson DI, Hughes D. Evolution of antibiotic resistance at non-lethal drug concentrations. *Drug Resist Up* 2012;**15**:162−72.
112. Gullberg E, Albrecht LM, Karlsson C, Sandegren L, Andersson DI. Selection of a multidrug resistance plasmid by sublethal levels of antibiotics and heavy metals. *mBio* 2014;**5**:e01918−14.
113. Baquero F, Negri MC. Challenges: selective compartments for resistant microorganisms in antibiotic gradients. *Bioessays* 1997;**19**:731−6.
114. Drlica K, Zhao X. Mutant selection window hypothesis updated. *Clin Infect Dis* 2007;**44**: 681−8.
115. Kimura M. Evolutionary rate at the molecular level. *Nature* 1968;**217**:624−6.
116. Kimura M. *The neutral theory of molecular evolution*. Cambridge University Press; 1983.
117. Ohta T. The nearly neutral theory of molecular evolution. *Ann Rev Ecol Syst* 1992;**21**: 263−86.
118. Hughes AL. Evidence for abundant slightly deleterious polymorphisms in bacterial populations. *Genetics* 2005;**169**:533−8.
119. Comas I, Chakravarti J, Small PM, Galagan J, Niemann S, Kremer K, et al. Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat Genet* 2010;**42**:498−503.
120. Austin DJ, Kristinsson KG, Anderson RM. The relationship between the volume of antimicrobial consumption in human communities and the frequency of resistance. *Proc Natl Acad Sci USA* 1999;**96**:1152−6.
121. Seppälä H, Klaukka T, Vuopio-Varkila J, Muotiala A, Helenius H, Lager K, et al. The effect of changes in the consumption of macrolide antibiotics on erythromycin resistance in group A streptococci in Finland. *N. Engl J Med* 1997;**337**:441−6.
122. Guillemot D, Varon E, Bernede C, Weber P, Henriet L, Simon S, et al. Reduction of antibiotic use in the community reduces the rate of colonization with penicillin G-nonsusceptible *Streptococcus pneumoniae*. *Clin Infect Dis* 2005;**41**:930−8.
123. Andersson DI, Hughes D. Antibiotic resistance and its cost: is it possible to reverse resistance? *Nat Rev Micro* 2010;**8**:260−71.
124. Maisnier-Patin S, Andersson DI. Adaptation to the deleterious effects of antimicrobial drug resistance mutations by compensatory evolution. *Res Microbiol* 2004; **155**:360−9.
125. Marcusson LL, Frimodt-Møller N, Hughes D. Interplay in the selection of fluoroquinolone resistance and bacterial fitness. *PLoS Pathog* 2009;**5**:e1000541.
126. Trindade S, Sousa A, Xavier KB, Dionisio F, Ferreira MG, Gordo I. Positive epistasis drives the acquisition of multidrug resistance. *PLoS Genet* 2009;**5**:e1000578.
127. Gagneux S, Burgos MV, DeRiemer K, Enciso A, Muñoz S, Hopewell PC, et al. Impact of bacterial genetics on the transmission of isoniazid-resistant *Mycobacterium tuberculosis*. *PLoS Pathog* 2006;**2**:e61.

Downloaded From Vet-eBooks.com

128. Sherman DR, Mdluli K, Hickey MJ, Arain TM, Morris SL, Barry CE, et al. Compensatory *ahpC* gene expression in isoniazid-resistant *Mycobacterium tuberculosis*. *Science* 1996; **272**:1641−3.
129. Ando H, Kondo Y, Suetake T, Toyota E, Kato S, Mori T, et al. Identification of *katG* mutations associated with high-level isoniazid resistance in *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* 2010;**54**:1793−9.
130. Borrell S, Gagneux S. Infectiousness, reproductive fitness and evolution of drug-resistant *Mycobacterium tuberculosis*. *Int J Tuberc Lung Dis* 2009;**13**:1456−66.
131. Olivares J, Álvarez-Ortega C, Linares JF, Rojo F, Köhler T, Martínez JL. Overproduction of the multidrug efflux pump MexEF-OprN does not impair *Pseudomonas aeruginosa* fitness in competition tests, but produces specific changes in bacterial regulatory networks. *Environ Microbiol* 2012;**14**:1968−81.
132. Sandegren L, Andersson DI. Bacterial gene amplification: implications for the evolution of antibiotic resistance. *Nat Rev Micro* 2009;**7**:578−88.
133. Nilsson AI, Zorzet A, Kanth A, Dahlström S, Berg OG, Andersson DI. Reducing the fitness cost of antibiotic resistance by amplification of initiator tRNA genes. *Proc Natl Acad Sci USA* 2006;**103**:6976−81.
134. Elena SF, Lenski RE. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet* 2003;**4**:457−69.
135. Campbell EA, Korzheva N, Mustaev A, Murakami K, Nair S, Goldfarb A, et al. Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell* 2001;**104**:901−12.
136. Ezekiel DH, Hutchins JE. Mutations affecting RNA polymerase associated with rifampicin resistance in *Escherichia coli*. *Nature* 1968;**220**:276−7.
137. O'Sullivan DM, McHugh TD, Gillespie SH. Analysis of *rpoB* and *pncA* mutations in the published literature: an insight into the role of oxidative stress in *Mycobacterium tuberculosis* evolution? *J Antimicrob Chemother* 2005;**55**:674−9.
138. Gagneux S, Long CD, Small PM, Van T, Schoolnik GK, Bohannan BJM. The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. *Science* 2006;**312**:1944−6.
139. Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, et al. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat Genet* 2012;**44**:106−10.
140. de Vos M, Müller B, Borrell S, Black PA, van Helden PD, Warren RM, et al. Putative compensatory mutations in the *rpoC* gene of rifampin-resistant *Mycobacterium tuberculosis* are associated with ongoing transmission. *Antimicrob Agents Chemother* 2013;**57**:827−32.
141. Qj L, Jiao WW, Qq Y, Xu F, Jq L, Sun L, et al. Compensatory mutations of rifampicin resistance are associated with transmission of multidrug resistant *Mycobacterium tuberculosis* Beijing genotype strains in China. *Antimicrob Agents Chemother* 2016. http://dx.doi.org/10.1128/AAC.02358-15.
142. Cohen KA, Abeel T, Manson McGuire A, Desjardins CA, Munsamy V, Shea TP, et al. Evolution of extensively drug-resistant tuberculosis over four decades: whole genome sequencing and dating analysis of *Mycobacterium tuberculosis* isolates from KwaZulu-Natal. *PLoS Med* 2015;**12**:e1001880.
143. Casali N, Nikolayevskyy V, Balabanova Y, Harris SR, Ignatyeva O, Kontsevaya I, et al. Evolution and transmission of drug resistant tuberculosis in a Russian population. *Nat Genet* 2014;**46**:279−86.
144. Price GR. Selection and covariance. *Nature* 1970;**227**:520−1.
145. Baquero F. Evaluation of risks and benefits of consumption of antibiotics: from individual to public health. In: Tibayrenc M, editor. *Encyclopedia of infectious diseases*. Hoboken, New Jersey, USA: Wiley and Sons, Inc.; 2007. p. 509−20.

Downloaded From Vet-eBooks.com

Modern Morphometrics of Medically Important Arthropods

13

J.-P. Dujardin
CIRAD-IRD, Baillarguet, France

1. Introduction

The phenotype is the product of the interaction between genes and environment. Phenotypic variation is then an expected outcome of more than one factor. It can be scored by measurable changes in anatomy, morphology, physiology, life history, behavior, and so on.[1,2] This chapter considers the phenotype as a set of metric properties and their variation.

1.1 Modern and Traditional Morphometrics

Morphometric techniques aim at measuring size, shape, and the relation between size and shape (allometry). Before the so-called "morphometric revolution,"[3] shape was an abstraction, a residue after scaling for size, and it was not possible to visualize this "residue." The replacement of initial variables describing a distance between two anatomical points by the coordinates of these points, and the subsequent visualizing techniques, represented a giant step in the direct study of forms.

For shape comparisons, great importance is given to the quality of landmarks in terms of comparability. Two conceptually and statistically separate approaches are: (1) landmark-based morphometrics, using the relative position of a few anatomical landmarks, and (2) outline-based morphometrics,[3,4] which captures the contour of forms through a sequence of close pseudo-landmarks.

2. Landmark-Based Geometric Morphometry

In common practice, size and shape are derived from a configuration of landmarks collected on a nonarticulated part, often a single organ (but see Ref. 5). The choice of suitable landmarks relies on their operational homology. In the morphometrics practice, homology is "correspondence of parts" with no specification about whether the parts correspond with respect to structure, development, or phylogeny.[6] If individuals belong to a single species, homologous landmarks are probably similar due to common descent because all members of the species come from a common ancestor. If they belong to different species, there is no guaranty that homologous landmarks are similar due to common descent, except if they are known to be descending from a common

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00013-5
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

ancestor.[7] This homology is one of the criteria making landmark-based morphometrics a suitable tool for systematics (see Section 10.1).

Anatomical landmarks are relocatable points, and according to this criterion various levels of quality have been recognized (see type I, II, and III landmarks, having decreasing levels of precision[8]).

2.1 Landmark-Based Size

Traditional systematists often selected one single measurement, for instance, the length of the wing along its largest axis, as an estimator of an insect body size.[9–11] Such relationship is often assumed rather than demonstrated.[10,12]

2.1.1 Size Variable: The Centroid Size

The centroid size (CS) is the square root of the sum of the squared distances from the centroid to each landmark (see Gower, 1971 in Ref. 13). Depending on the relative position of all landmarks, this measure is the most inclusive one. It has been shown that, in the case of small, circular variation at each landmark, this estimator of isometric change of size is not correlated to shape variation.[8]

The relationship of CS values and the traditional wing length in the mosquito *Aedes aegypti* showed good correlation.[12] Actually, the correlation of CS values with traditional inter-landmark distances (ILD) is itself correlated to the relative dimensions of ILD: the largest the ILD, the highest its correlation with CS (Fig. 13.1).

2.2 Landmark-Based Shape

In many fields where morphometrics is applied, shape has been traditionally described as the ratio of one dimension to another. Although intuitively the ratio may appear as capable of scaling for size, it often does not.[14–17] Moreover, the ratios introduce some well-known statistical drawbacks.[15] Angles do not improve the situation since they are another kind of ratio.[16] In geometric morphometrics, the shape of a configuration of landmarks is represented by their relative positions as contained in their coordinates.

2.2.1 Procrustes Residuals, Partial Warps, Relative Warps, and Tangent Space Variables

The raw landmarks' coordinates also contain artifactual variation due to position, size, and orientation. Shape must be described by new variables having removed these artifacts. This is obtained through the *Procrustes* superimposition on a consensus configuration. If using the least squares fit as an optimality criterion, the statistical procedure of superposition is called generalized Procrustes analysis (GPA). It is currently the most common procedure, but other techniques also exist.[18] The residual coordinates after a GPA depend on the composition of the group under study. If other specimens

Downloaded From Vet-eBooks.com

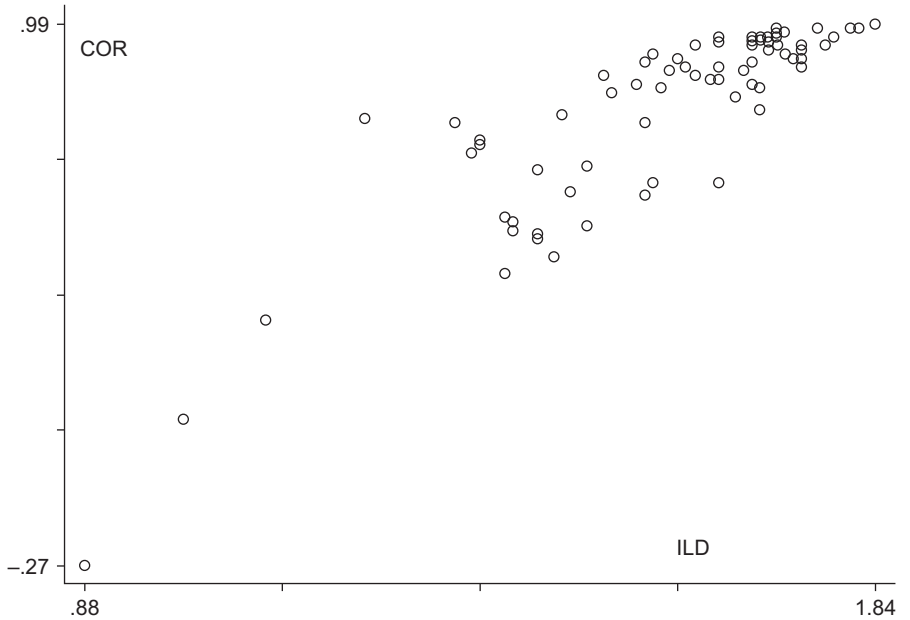


Figure 13.1 Each circle represents, on the vertical axis, the correlation found between an inter-landmark distances (ILD) and the centroid size (CS) of the wing computed from the total set of landmarks. The value of each ILD on the horizontal axis is an average (mean ILD) obtained from the total number of wings (78) examined in this sample. The wings belong to *Aedes aegypti*.

(i.e., coordinates) are added to the analysis, shape variables must be recomputed accordingly.[3,19]

The superimposition procedure induces the loss of four degrees of freedom[20]: the residual coordinates must be further modified by a rigid rotation so that they can be studied using classical statistical tools.[21] Thus, the aligned configurations (Procrustes coordinates) are projected onto an Euclidean space that is tangent to the curved space (shape space) at the consensus configuration. The coordinates in the tangent space are then used as shape variables ("tangent space variables," or "Procrustes residuals") in all subsequent analyses. An alternative, mathematically equivalent procedure consists of using eigenvectors of the bending-energy matrix.[8] These so-called "warps methods" tend to become less commonly used than the direct analysis of the coordinates.

The resulting shape variables are generally submitted to a principal component analysis, and the principal components become the final shape variables.[22]

2.3 Semilandmarks for Curves Shape

A special development of type III landmarks, called "semilandmarks," allows the description of curved lines between two classical landmarks.[23–27] The development of semilandmark-based analysis allows one to use the "Procrustes paradigm" as a

Downloaded From Vet-eBooks.com

unified framework for both outline and landmark-based analyses. It does not make the outline approach an obsolete one, since the latter may apply to structures without any anatomical landmark (see Section 3). Currently, the methods specifically developed for semilandmarks make use of a minimization criterion,[28] either minimum bending-energy[8] or minimum Procrustes distance.[29] As they imply the sliding of points along the tangent to the curve (or tangent plane to the surface), the term "sliding landmarks" is also used to refer to semilandmarks. Both methods for semilandmarks processing do not produce exactly the same results.[28,30]

3. Pseudo-Landmark-Based Shape

Anatomical, "true" landmarks are opposed to "pseudo-landmarks" used in the outline-based approach. Pseudo-landmarks describe contours or boundary outlines and can exist with no anatomical landmark at all, or can include one or more of them. Pseudo-landmarks are of another nature than true landmarks because comparability is not expected from them separately, but from the structure they describe. Thus, if carefully chosen, like the contour of a tick's body,[31] the genital leaflet of a mosquito,[32] or an internal cell of the wing,[33,34] a contour represents a homologous structure allowing interindividual and/or intergroup comparisons.

3.1 Outline-Based Shape

Various techniques allow to derive size and shape variables from a digitized outline.[13,35–37] We give few words here about the most common one, the elliptic Fourier algorithm (EFA).

Briefly, the observed contour is decomposed in terms of sine and cosine curves of successive frequencies called harmonics, and each harmonic is described by four coefficients. With this method,[35] the first harmonic ellipse parameters are used to standardize the Fourier coefficients so that they are invariant to size, rotation, and the starting position of the outline trace. By doing this, the three first coefficients become constant (1, 0, and 0) and are not used in the remaining analyses. The fourth coefficient is related to the width-on-length ratio of the outline.[38–40]

Ideally, more harmonics should be able to capture more shape parameters and produce a higher level of discrimination. However, more harmonics also inflate the digitization error, and an optimal number should be selected.[41,42]

3.2 Outline-Based Size

The Fourier's coefficients described earlier are normalized by the semigrand axe of the first harmonic ellipse. The latter may be used as an estimate of global size, as well as the root-squared area of this same starting ellipse,[13] the root-squared area of the digitized contour, or its perimeter.

Downloaded From Vet-eBooks.com

4. Allometry

The study of allometry has a long history in biology,[43] which has mostly focused on the covariation among structures: for example, the relationship between fore- and hind limbs in theropod dinosaurs (e.g., *Tyrannosaurus rex*), or between brain size and body height in humans.

The relationship between size and shape is called allometry. Various kinds of allometry have been described, which might result from the different developmental processes. During development, organisms grow larger and in the meantime, because the rate of growth differs between their various parts, their shape also changes. Such covariation between size and shape throughout the ontogeny is called "ontogenetic" allometry. When considering different individuals of the same population and at the same developmental stage, there is also often some variation in size and shape. Such allometry is referred to as "static" allometry and is opposed to the "evolutionary" allometry that can be identified across species or geographic populations.

Geometric shape variables (see previous paragraph) are not allometry-free variables, there is an allometric residue contained in the shape variables. Geometric morphometrics, by clearly separating shape and size, allows one to specifically investigate their relationship. In some circumstances, one might be tempted to remove the allometric residue of shape variation. This could be justified when dealing with the "static" allometry, to tentatively remove the environmental influence on the metric properties of conspecific populations.[12,14,44] The tentative removal of the allometric effect on shape seems less justified for the "evolutionary allometry." For interspecific comparisons, indeed, the allometric variation is likely to be part of the evolutionary process, its inclusion in the variable is relevant to systematics.

5. Measurement Error

The measurement error exists at various steps of morphometric analysis.[45] The mounting technique of specimens or organs, the photographing conditions, and the user's skill to collect landmark coordinates may produce artifactual variation.

It is convenient to photograph a sight (chart) to make sure there is no distortion of the image, which should be placed in the center of the visual field to avoid possible optical defects at the periphery of the lens.

Whatever the quality and reproducibility of landmark digitization, the recommended way to perform morphometric comparisons is to allow one single user to produce the data.

Even when performed by a single user, digitization should be repeated at least once, allowing one to measure the precision[45,46] and to reduce the error by averaging repeated measures.

Downloaded From Vet-eBooks.com

6. Some Considerations About the Genetics of Metric Change

Does shape have a strong genetic determinism or is it mostly plastic? How many genes are involved in the shape variation? Are there few genes with major effects or many genes of small effects?

6.1 Shape As a Polygenic Character

Shape appears as a classical polygenic character.[47] Evidence for strong genetic determinism of shape was suggested by significant association with chromosome polymorphism,[48–50] and confirmed by genetic studies.[51,52] When studies on quantitative trait loci (QTL) were applied to the shape and size of mouse mandible, many QTL were identified for shape,[53] many more than for size.[54,55] Few studies are found in insects, also fitting the idea of genetic determinism[56] and polygenic inheritance.[57,58]

6.2 Genetic Drift

Since landmark-based or outline-based shape seems the output of a cascade of genes, it is expected that in natural conditions genetic drift be a common factor of shape variation. Field observation frequently reported significant shape differences between geographic populations.[59–65] Using a set of three isofemale lines of *A. aegypti* monitored during 10 generations, a significant shift of shape appeared in one line, with nonsignificant changes in corresponding size.[66] In this experiment, the change apparently produced by genetic drift did not affect the same landmarks as those affected by larval food or density variation.[67]

6.3 Heritability

Size in insects may show consistent heritability values,[11,68] so that they can be experimentally selected to constitute subpopulations genetically distinct for size.[69,70]

In quantitative genetics, the response to selection is embodied in the breeders' equation R = h2s,[71] where "s" is the selection differential, "R" is the response to selection, and "h2" is heritability. Heritability ("h2") is a ratio, it is the fraction of phenotypic variance (VP) that is due to genetic differences (VG), it plays a central role in this equation and conditions the response to selection (see Ref. 72 for a critic of the use of heritability).

The problem that arises here is that shape is inherently a multivariate trait. As such, its genetic variation cannot be assessed by a scalar heritability as for univariate quantitative traits. While it is theoretically possible to derive such a univariate measure of shape genetic variation using some kind of shape distance (e.g., Procrustes distance, see Ref. 73), it has been shown to be misleading.[74,75] For example, imagine a situation where such a global shape heritability would be high say, 0.8. This would give the false

Downloaded From Vet-eBooks.com

impression that shape could be selected in any direction. However, the multivariate version of the breeder's equation[76] shows that the multivariate response to selection is highly dependent on the genetic correlations among traits (the G matrix). For shape, G represents the genetic variances and covariances among landmarks coordinates, or shape dimensions. And these correlations are by no means uniform in the shape space. In other words, some combinations of shape changes are generally genetically more variable than others, imposing selection a relative constraint. It is even conceivable that some directions are fully devoid of any variation at all, a situation referred to as an absolute constraint.[77] This means that even with a high value of global heritability, because of the anisotropic nature of genetic variation (i.e., genetic variation is not evenly distributed in the shape space), it might well be impossible to select at all in some specific direction (i.e., some shape changes are impossible to reach through selection).

These considerations make the analysis of the evolution of G and P (phenetic correlations among traits) very important to understand the constraints and potential of shape evolution.[78–80] This relates to the general field of modularity/integration that investigates the patterns of covariation among parts, and a substantial amount of literature using geometric morphometrics has been devoted to it.[79,81]

In insects, morphological traits commonly have the highest heritability values compared to other types of traits, such as life history, probably because the former are less concerned with fitness.[82] The consistent values of shape heritability measured so far[48,82–84] suggest that a large fraction of morphometric divergence seen between natural populations of insects[12,63,65] may be due to additive effects of genes.

Since heritability is a measure of genetic variation, it depends on the population under study. In population studies related to medically important arthropods, the measurement of heritability is not mandatory to the epidemiological interpretation of natural metric variation (see Section 10.5.2).

6.4 Hidden Genetic Variability

Unexpectedly, some heritable changes seem to be triggered by environmental events. Such phenomenon, which is reminiscent of Lamarkian "inheritance of acquired characteristics," has been named "genetic assimilation"[85] or "autonomization" (Schmalhausen, 1949 in Ref. 86), and, later, "genetic accommodation."[1]

The mechanisms by which an environmentally induced phenotype may become heritable are entirely compatible with concepts of classical neo-Darwinian evolutionary biology. Indeed, the environmental trigger (since this is the disputable one) just uncovers previously cryptic genetic variation.[87] Thus, there are genetic mutations that can remain masked until the environment (or another mutation) reveals them.[88] Adaptationist ("capacitator," see for instance, [89–92]) and other theories ("robustness" as a generic property of complex systems, see for instance, [93–95]) provided plausible scenarios explaining the existence of hidden genetic variability.

Downloaded From Vet-eBooks.com

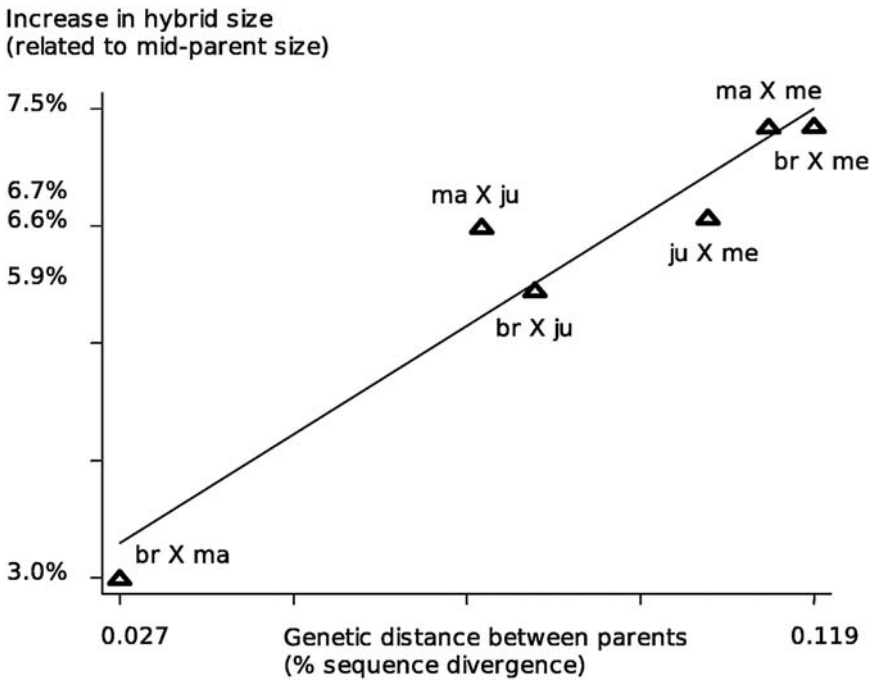


Figure 13.2 Relationship between size in hybrids and genetic distances between parents belonging to the Braziliensis complex. On vertical axis, the increasing of size relative to mid-parent size. On horizontal axis, the genetic distance between parents as inferred from mitochondrial DNA sequence. *br*, *T. b. brasiliensis*; *ju*, *T. b. juazeirensis*; *ma*, *Triatoma brasiliensis macromelasoma*; *me*, *T. b. melanica*.
From Dujardin JP, Costa J, Bustamante D, Jaramillo N, Catalá S. Deciphering morphology in Triatominae: the evolutionary signals. Acta Trop 2009;**110**:101−11.

6.5 Hybridism

Various observations have been made in the vectors of Chagas disease, where interspecific hybridism is relatively common.[96]

In the brasiliensis complex (Triatominae), historical hybridism has been suggested as a possible homoploid speciation mechanism for one of its members, *Triatoma macromelasoma*, which would have arisen from the hybrid cross between *Triatoma juazeirensis* and *Triatoma brasiliensis*.[97,98] Today, it is still possible to cross *T. brasiliensis* and *T. juazeirensis*,[99] as well as to cross among the four members of the Brasiliensis complex. An experimental study disclosed a linear relationship between the genetic divergence of the parents[99,100] and the increase in size of their offspring (Fig. 13.2). Contrary to the size, the shape of the hybrids remained intermediate between parents.[97]

Between cryptic species of *Diachasmimorpha longicaudata*, a hymenopteran parasitoid of fruit flies, the size of the hybrids was larger than that of mid-parents, although not significantly larger,[102] and hybrids showed an intermediate landmark-based shape.

Downloaded From Vet-eBooks.com

These two studies suggested different genetic mechanisms affecting size and shape. Interestingly, size appeared as a character prone to show heterosis in case of genetically differentiated parents. No such heterosis was observed for the progeny between seven laboratory colonies of *Triatoma protracta*,[103] or between each of the five subspecies of *T. protracta*, suggesting low genetic divergence between them.

7. Phenotypic Plasticity

The genotype does not give rise to a single phenotype, but to a range of possible phenotypes. The "reaction norm" is the whole repertoire of possible phenotypes that may occur for a given genotype in all environments.[104,105] The reaction norm can easily be explored in laboratory experiments.[44,67,106–108] By definition, phenotypic plasticity (PP) is the occurrence of phenotypic variation of a single genotype interacting with different environments.[105]

A new phenotype expressed in a new environment may be adaptive. To this condition, PP can aid speciation by making available a different phenotype upon which natural selection can act. In such scenario, speciation would start with PP, not reproductive isolation.[109] In Triatominae, such a scenario is apparent,[110] and many examples exist of morphologically and ecologically recognized species that can still interbreed.[101] Understanding the causes and consequences of phenotypic variation is important for understanding the mechanisms of evolution. However, the genetic mechanisms underlying the evolutionary importance of PP[1,85,105,111] received so far only few experimental confirmations.[89,91]

Contrary to its evolutionary importance, the ecological importance of PP is easy to understand: populations or species having wider adaptive plastic responses can enlarge their ecological niches. For instance, among the more than 140 species of Triatominae, a few species have been able to colonize human structures. Within some of these species, the comparison of "domestic" and sylvatic subpopulations highlighted significant size differences, sylvatic insects being generally larger. Were these species more plastic than the others so that they could reduce their size as apparently required by survival in artificial ecotopes? Was the size a secondary event selected by the domestic environment? In *Rhodnius pallescens*, the significant size reduction was experimentally shown to be a plastic response to combined population density and feeding frequency parameters.[44]

In addition to explaining diversity and adaptation, PP also impacts our understanding of taxonomy, because it suggests that species characteristics are not immutable, but are influenced by the environment and can be highly variable.[112]

8. A Special Case of Shape Change: the Character Displacement

The initial definition of the "character displacement" (CD) concept[113,114] did not predicate the real complexity of its demonstration. The difficulties of obtaining

Downloaded From Vet-eBooks.com

unambiguous evidence from natural observations have been discussed by Ref. 115, and, subsequently, by.Ref. 116 Typically, CD was suspected when more difference was observed between species developing in sympatry than in allopatry.[116] Various more conditions must be satisfied to assess CD, among which the level of differences in sympatry (greater than expected by chance), the evolutionary history of sympatry (original or derived situation), the genetic nature of phenotypic differences, and, importantly, the connection between characters and competition for resources.[117] CD was demonstrated for behavioral and ecological characters more often than for morphological characters.[115,116] Morphologically, the displaced character is expected to be part of the feeding apparatus.[118] If the mouth-parts have a species recognition function, then displacement may have consequences on speciation as well.

9. The Regulation of Phenotype

To the many sources of phenotypic changes, the organism opposes homeostatic processes. Two components of this homeostasis are "canalization" and "developmental stability (DS)." The two components may be difficult to define: canalization would be the between-individual stability of development, DS would refer to the within-individual stability. Canalization may be defined as a buffering process against external and/or mutational perturbation from one (micro)environment to another,[119] while DS allows the organism to withstand random accidents during development in the same environment.[120]

For a review of these concepts and their many definitions, please see Ref. 121. Canalization and DS have been suggested to be independent processes,[122,123] but this position is not consensual, and others have suggested that canalization and DS are the same thing,[51] or even that there is no need for any specific buffering process.[124]

9.1 Canalization

The term "canalization" is due to Waddington, corresponding to the "stabilizing selection" of Schmalhausen.[86]

As for PP, canalization is not a property of a species or of a population, but of a genotype.[125] However, different traits of a single organism can be examined for their relative canalization by studying their natural variation in different lines, populations, or species. For instance, contrary to size changes, shape changes of the wings induced by striking altitudinal variation as found between the Andes and the Amazon basin could not interfere with species differences in sandflies.[61] A similar study comparing the wing shape of transcontinental populations of two close mosquito species, *A. aegypti* and *Aedes albopictus*, showed that species differentiation based on wing shape, but not on its size, was not altered by transcontinental migration during the last decades. Both species were still distinguishable at the same landmark locations.[65] This relative constancy of shape patterns within

Downloaded From Vet-eBooks.com

each species contrasted with the lability of size. For the same species (*A. aegypti*), size was significantly affected by a simple change in the food concentration or in the larval density.[67] Another example comparing size and shape responses is found in highly inbred lines of *R. pallescens* (Triatominae): the plastic response scored for the CS of the wing to the laboratory conditions of "domesticity" was not observed for the shape, except as an allometric change.[44] The apparently higher canalization of shape makes this trait a suitable character for populations and species distinction.[62]

9.2 Developmental Stability

Stress, be it environmental or genetic, and other also causes, such as relaxation of selection pressure (domestic animals), may have a disturbing effect on the organism development. The use of morphometrics as an indicator of stable development has been performed by estimating the frequency of abnormal phenotypes (phenodeviants). More effectively, it has been estimated by the amount of fluctuating asymmetry (FA).[126]

FA of size is a nondirectional asymmetry, as opposed to the classic directional asymmetry where one side is, on average, larger from the other side (which is common for internal organs of vertebrates). For a given sample or population, FA is computed as the variance of bilateral differences scored between individuals, either of size or of shape.[46,126] It has to be distinguished from antisymmetry, another kind of nondirectional asymmetry, less frequent however, where signed bilateral differences do not have a Gaussian distribution.

Stress is the most commonly assigned cause to FA; it can be an infection by a virus or a parasite, a difficult conquest of a new habitat,[110] or simply a different source of food.[127] In *Glossina* flies (vectors of sleeping sickness and nagana), five species were compared between laboratory and natural conditions of life, showing a strikingly reduced FA in the laboratory (Kaba et al., unpublished data). In natural conditions, the adults caught during the rainy season showed significantly higher FA than those collected in the dry season (Djohan et al., Ph.D. thesis at the H-B University of Abidjan).

Although FA has been widely used as a bioindicator of environmental and genetic stress, quite a strong controversy has occurred in the late 1990s, mostly around its genetic bases and in particular its heritability.[128] It has been suggested that publication biases have occurred, casting doubts upon the reliability of the literature on FA and in particular its role as an indicator of fitness.[129,130] One should thus use FA very cautiously when attempting to relate it to any other population parameter.

Other measures of DS could be used that are considered as developmental invariants, such as fractal dimensions, although they were described for plants and vertebrates only.[120] The use of shape variation itself in response to environmental stress has been advocated for insects.[131] However, contrary to symmetry that is expected to be perfect, and contrary to the frequency of phenodeviants that is expected to be zero, there is no "expected shape" in case of undisturbed development.

Downloaded From Vet-eBooks.com

10. Applications in Medical Entomology

10.1 Species Identification and Detection

The most important objection to the morphological concept of species is the existence of sibling (or isomorphic) species.[132] Sibling (or also cryptic) species are morphologically identical or nearly identical entities in spite of being recognized as different species according to other, modern concept(s) of species. The modern concepts of species make use of other criteria than simple morphological comparison, with some of them even completely free of any character examination. The most frequently used modern species concept in entomology is the biological one,[132,133] but with the increasingly used molecular techniques, other concepts are invoked, such as the Hennigian concept,[134] the evolutionary concept,[135,136] and the various phylogenetic concepts.[137]

However, this objection to the typological concept (i.e., to "morphospecies") is weakened by the possibilities of modern quantitative shape comparisons.[138–141] Shape comparisons detect minimal morphological variations, which often are undetectable by traditional morphological studies and even by classical morphometric approaches. Cryptic species of insects showed distinct shapes of wing venation in kissing bugs,[101,142,143] sandflies,[59] mosquitoes,[144,145] scythrididis,[146] parasitoid hymenoptera,[102,138,147] syrphids,[148] fruit flies,[149] and screwworm flies.[150] Although morphometric discrimination does not necessarily mean species determination, it has also been used to question species boundaries,[64] or to synonymize controversial taxa.[60]

Because landmark-based shape is defined relative to the consensus of the specimens under study, shape variables derived from one set of coordinates cannot be compared with shape variables derived from another set. Coordinates themselves could be used for such comparisons, but the measurement error may represent a significant obstacle, especially when the objective is to distinguish very similar species.

The User Effect. The error due to the user is generally due to small but systematic differences in pointing to the exact localization of some landmarks. These subtle discrepancies are amplified by the power of multivariate analysis, such as the discriminant analysis.[151] Their impact can be reduced averaging repeated collections of the data.[45] However, such correction might not be satisfactory when comparing very close specimens or groups, and measurement error may become a significant obstacle for different users.[152,153]

The Need for a Bank of Reference Images. To circumvent the lack of exchangeability of the morphometric variables, an alternative geometric descriptive system should be developed that separates data gathering and analyses. It goes through the creation of a bank of reference images from which one can extract raw data and compare it to external, unknown specimens.[151] Such an initiative is ongoing at http://mome-clic.com under the name CLIC (Collection of Landmarks for Identification and Characterization); two other such banks are currently developed, one for mosquitoes (http://wingbank.com.br/), and one for bees (http://apiclass.mnhn.fr/).

Downloaded From Vet-eBooks.com

10.2 Characterization Tool at the Individual Level

Using geometric shape comparisons, one single individual can generally be accurately classified using a database of images of the candidate species.[143] As an example, we show here unpublished data about mosquito identification. Each single individual has been allocated to its closest group (according to Mahalanobis distance) without using that individual to help determine a group center ("validated reclassification"). The wing venation patterns of Culicidae, in spite of being roughly the same among the genera, allowed an almost perfect reclassification (Table 13.1). Within some genera, such as *Aedes* or *Anopheles*, the species discrimination was very satisfactory; it was less convincing in the genus *Culex* (Table 13.2). The possibility to perform satisfactory identifications without being an expert in taxonomy is very attractive, and encouraging results have been obtained for tsetse flies classification (Kaba et al., unpublished data), but more studies are needed to evaluate the full interest of this identification approach in many groups of medically important insects.

More challenging is the comparison between conspecific individuals. Reinfestant specimens after vector control measure may be few, and classical morphology could be unable to suggest their origin (see Section 10.4). Provided a database exists on specimens collected before control measures, shape can be used for quantitative comparisons of local and external individuals.[103]

10.3 Biodiversity

The transmission of vector-borne diseases has obvious links with the environment. Studies exploring these links suggested that the reduction in global biodiversity is likely to contribute to vector-borne disease transmission through the "dilution effect."[154,155] It is therefore highly desirable to quantify the environment. In this kind of study, geometric morphometrics has two advantages to offer: one is its ability

Table 13.1 Morphometric Identification of Culicidae Based on 13 Landmarks of the Wing

Genera	Ur, Ma	An	Mi	Cu	Ae, Ar, Co
Scores (%)	100	97	96	95	100
N	508(8)	446(6)	348(5)	317(4)	127(3)

The first column indicates that 100% of the genus *Uranotaenia* (Ur) and 100% of the genus *Mansonia* (Ma) could be recognized when mixed with the six other genera: *Anopheles* (An), *Mimomyia* (Mi), *Culex* (Cu), *Aedes* (Ae), *Armigeres* (Ar), and *Coquillettidia* (Co). The second column indicates that 97% of the genus *Anopheles* could be recognized when mixed with the genera *Mimomyia*, *Culex*, *Aedes*, and *Armigeres*. The third column indicates that 96% of the genus *Mimomyia* could be recognized when mixed with the genera *Culex*, *Aedes*, and *Armigeres*. 95% of the *Culex* could be distinguished from the genera *Aedes*, *Armigeres*, and *Coquillettidia*. The last column indicates that these three genera were perfectly discriminated by their wing geometry. *Between brackets*, number of genera in the analysis; *N*, total number of individuals in each analysis.
Mosquito collection by Henry A and Thongsripong P (University of Hawaii). Morphological identification of the genera by Dr. Rattanarithikul R (AFRIMS, Thailand). Digitization of wings by Lasnes J-F (University of Montpellier).

Downloaded From Vet-eBooks.com

Table 13.2 Correct Species Attribution Scores Based on the Geometry of the Wings

Species	Scores (%)	n/N
Aedes		
(Stegomyia) aegypti	100	12/12
(Neomelaniconion) lineatopennis	66	10/15
(Aedimorphus) mediolineatus	100	12/12
(Aedimorphus) vexans	83	20/24
Anopheles		
(Anopheles) barbirostris	100	14/14
(Cellia) tesselatus	88	8/9
(Cellia) vagus	91	34/37
Culex		
(Culex) vishnui	55	29/52
(Culex) gelidus	61	11/18
(Culex) quinquefasciatus	91	11/12
(Oculeomyia) bitaeniorynchus	78	18/23
(Oculeomyia) sinensis	62	18/29
(Culiciomyia) nigropunctatus	91	11/12

Thirteen species belonging to three genera, *Anopheles* (An.), *Culex* (Cx.), and *Aedes* (Ae.), were analyzed for species identification, namely: *Ae. aegypti*, *Ae. lineatopennis*, *Ae. mediolineatus*, *Ae. vexans*, *An. barbirostris*, *An. tessellatus*, *An. vagus*, *Cx. bitaeniorhynchus*, *Cx. gelidus*, *Cx. nigropunctatus*, *Cx. quinquefasciatus*, *Cx. sinensis*, and *Cx. vishnui*. *N*, total number of individuals in the species; *n*, number of individuals correctly assigned to the species; *Scores*, correct attributions in percentages by species after validated reclassification.
Mosquito collection by Henry A and Thongsripong P (University of Hawaii). Species morphological identification by Dr. Rattanarithikul R (AFRIMS, Thailand). Digitization of wings by Lasnes J-F (University of Montpellier).

to help the identification of the taxa (see Tables 13.1 and 13.2), the other one is its own addition to the knowledge about biodiversity.

In addition to specific estimates of biodiversity, such as the Shannon−Wiener[156] and the Simpson[157] indexes, complementary information has been looked for in the morphological disparity of organisms. Modern morphometrics provides quantitative tool for accurate metric disparity (MD) measurements and comparisons.[158,159]

One could expect a higher richness to be the cause of higher morphometric variation; if selection targets form rather than species, some relationship is predictable.[160] However, no such relationship could be confirmed: trends in species richness (SR) generally did not match trends in MD.

Performing simulations from true data (43 mosquito species from six different environments), we observed that the correlation was always positive between MD

Downloaded From Vet-eBooks.com

and species richness was not satisfactory when using the full set of landmarks available. However, it could increase to very high values when a very limited set of LM was used.[161]

10.4 Reinfestation Studies

Reinfestation studies refer to the situation where insects that have been controlled by any method are coming back to the treated area. Are they migrants coming from neighboring untreated foci, or are they descendants of a residual population that escaped the applied control methods?

As long as geometric shape is able to identify the parental generation and to distinguish it satisfactorily from other subpopulations,[71,103] it might be able to provide relevant information in studies of reinfestation after treatment.[162,163]

Provided that samples were available from the population before insecticide application, relative similarities could suggest the origin of reinfesting specimens,[162] and such information has been shown to be in agreement with genetic markers.[163] The geometry of the wing (landmark-based shape) of *T. protracta* was tested on laboratory populations and was shown to be an interesting candidate to assess the origin of a given individual.[103]

The reinfestation analysis is based on the simple assumption that an insect is more similar to its parents than to other insects. Moreover, since the objective is to distinguish local "inhabitants" from "immigrants," the possible environmental effect on metric traits (the environmental covariance) is a welcome effect.

Of course, reinfestation studies would be less applicable to highly dispersive insects breaking the population structure at each new generation.[164,165]

10.5 Population Structure

A recurrent need in medical entomology is to quantify the current exchanges of individuals among subpopulations. This quantification would inform on "population structure," to be distinguished from "genetic" structure, which is defined by the level of gene flow among subpopulations. Although mark-release-recapture studies might be a valid option to evaluate the frequency of active migrants among subpopulations,[166,167] it cannot account for passive migration of nonflying stages of the insect, so that this frequency is currently evaluated by indirect methods; the measurement of gene flow is the technique of choice.[168,169]

10.5.1 Gene Flow and the Flow of Migrants

Gene flow measurement provides indirect information on the level of migration among subpopulations. Lack of gene flow is a valid information since in that circumstance (genetic divergence) migrants are highly unlikely. Less valid information is the similarity of gene frequencies, since no one can affirm that such lack of genetic structure is a reflection of the current level of migration. How contemporaneous or recent it depends on the effective size of the populations under study and on the evolutionary

Downloaded From Vet-eBooks.com

rate of the genetic marker.[170] Additional problems with genetic markers are that they are relatively costly and they need appropriate infrastructures. Genetic markers often remain inside research laboratories and have not yet found their way into routine medical entomology.

10.5.2 Environmental Variance of Size Versus of Shape

Modern morphometrics is tempting as a candidate population marker because it is a fast, low-cost, easily spread tool, it is informative about current or very recent population events,[71] and it contains information on genetic variation. However, as long as morphometric traits have much higher environmental variance than genetic markers, they are not appropriate for gene flow estimation.

How then to interpret geographic variation of metric properties? Metric variation can be decomposed into size and shape variation so that their environmental variance can be examined separately. The importance of diversifying selection inflating size or shape variation among natural populations can be quantified by comparing the same material (1) the *Fst* index as derived from neutral molecular markers and (2) the *Qst* index as computed from metric characters. *Qst* separates quantitative genetic variation in a manner analogous to *Fst* for single gene markers[171]: if the quantitative characters and the molecular characters are neutral, *Qst* and *Fst* should converge to the same value.[172–174] Data comparing molecular *Fst* and quantitative *Qst* are few. They tend to show the following trends: (1) *Qst* is generally higher, or much higher, than *Fst*, and (2) the value of *Qst* depends on character fitness.[170] Within species, traits experiencing the strongest local selection pressures (diverging, or diversifying selection) are expected to be the most divergent from molecular *Fst*.[170]

The small set of comparisons in medically important insects between *Qst* and *Fst* confirmed the importance of selection modifying the geometric variation among subpopulations.[140] These comparisons allowed two more observations: (1) in agreement with the idea of shape having less environmental variance than size, they confirmed the lower sensitivity of shape (relative to size) in response to diversifying selection, and (2) in agreement with the infrequent report of a *Qst* lower than *Fst*, which would suggest homogenizing selection acting on the quantitative trait, no such situation was observed in medically important insects.[140]

10.5.3 Biogeographical Islands

Local elimination of an insect vector of disease is generally held to be feasible only for geographically constrained situations, such as islands. The task of population genetics studies is, in a sense, to find and define those biogeographical "islands"[175] of the vector distribution on the mainland.

Can modern morphometrics help define these target areas?

A few studies on mosquitoes[176] or kissing bugs,[177,178] but mainly on tsetse flies,[63,179–181] brought encouraging arguments to use geometric morphometrics as a tool to examine population structure.

Downloaded From Vet-eBooks.com

To discuss this application, it is important again to insist on which metric property is considered, either geometric shape or size. Here, we focus on landmark-based shape. Since the populations compared are conspecific ones, and especially if size variation is important, allometry-free shape should be preferred to just shape.

Such shape similarity between natural populations would be hardly explained by homogenizing selection (see $Qst > Fst$, Section 10.5.2). In our point of view, shape similarity is maintained by genetic exchanges, otherwise it is likely to be quickly broken because of genetic drift.

Between truly isolated populations, differences in shape should develop because of two main reasons: (1) genetic drift is likely to be a major force affecting shape, and (2) homogenizing selection seems infrequent (or unable to counteract the effects of genetic drift). Some experimental and natural observations agree with these propositions. For instance, two isolated isofemale laboratory lines of *A. aegypti* could diverge in shape (the geometry of wing venation) after less than 15 generations, in spite of an identical laboratory environment.[66] Studies performed on natural populations of insects could show that shape similarity was suggestive of exchanges between compared populations, although they were collected in different habitats. Thus, between populations of houses and of palm trees, the geometry of the wing venation of *Rhodnius prolixus* did not show significant differences.[182] Such similarity strongly suggested exchange of individuals, thus gene flow, which was confirmed later by genetic markers.[183]

Lack of isolation was also described by both genetic and metric markers for tsetse flies along the Mouhoun river in Burkina Faso,[179] as well as in the city of Abidjan between three sites a few miles away from each other.[181] The complete lack of landmark-based shape difference between Japan and USA populations of *A. albopictus* could suggest ongoing (passive) exchanges between these countries.[176]

10.5.4 The Need for a Heuristic

Considering the cost represented by the molecular machinery in developing countries, the earlier examples suggest that a faster and less-expensive morphometric approach could be helpful, even as an orientation technique only. Thus, geometric shape variation could be our guide to quickly identify at low-cost areas where isolation is possible and where it is unlikely. Two directives helping interpretation could be the following:

- If landmark-based shape does not show differences between populations, the most likely explanation is that populations are not isolated ones.
- If landmark-based shape shows strong differences, one should consider also the habitats that are compared: isolation is a reliable interpretation in case of similar environments.

These guidelines are based on the hypothesis that genetic drift is the main force in nature producing fast differences in shape among conspecific isolated populations. They refer to contemporaneous time, not to an undefined evolutionary past. They are easy to falsify, so that they invite for more natural observations related to population structure and shape variation.

Downloaded From Vet-eBooks.com

Glossary

Procrustes Whose name means "he who stretches," was a thief in Greek mythology (the myth of Theseus). He preyed on travelers along the road to Athens. He offered his victims hospitality on a magical bed that would fit any guest. As soon as the guest lay down, Procrustes went to work upon him, either stretching the guest or cutting off his limbs to make him fit perfectly onto the bed (Grose Educational Media, 1997–98).

Partial warps, relative warps, and so on A complete glossary of the many technical terms related to geometric morphometrics can be found at http//life.bio.sunysb.edu/morph/.

Acknowledgments

I thank V. Debat (MNHN, Paris) for his kind help in the revision of this chapter, particularly about shape heritability issues. This study has been supported by IRD grants number HC3165-3R165-GABI-ENT2 and HC3165-3R165-NV00THA1.

References

1. West-Eberhard MJ. Phenotypic plasticity and the origins of diversity. *Annu Rev Ecol Syst* 1989;**20**:249–78.
2. Gadagkar R, Chandrashekara K. Behavioral diversity and its apportionment in a primitively eusocial wasp. In: Ananthakrishnan TN, Whitman D, editors. *Insect phenotypic plasticity diversity of responses*. USA: Science Publishers, Inc. Enfield (NH); 2005. p. 108–24. p. 213.
3. Rohlf FJ, Marcus LF. A revolution in morphometrics. *Trends Ecol Evol* 1993;**8**(4): 129–32.
4. Lestrel PE. *Morphometrics for life sciences*. World Scientific Publishing; 2000.
5. Adams DC. Methods for shape analysis of landmark data from articulated structures. *Evol Ecol Res* 1999;**1**(8):959–70.
6. Smith GR. Homology in morphometrics and phylogenetics [Special Publiation Number 2]. In: Rohlf FJ, Bookstein FL, editors. *Proceedings of the Michigan morphometrics workshop*. Ann Arbor, MI: The University of Michigan Museum of Zoology; 1990. p. 325–38. p. 380.
7. Lele SR, Richtsmeyer J. *An invariant approach to statistical analysis of shape*. Boca Raton: Chapman and Hall/CRC; 2001. pp. VIII + 308.
8. Bookstein FL. *Morphometric tools for landmark data. Geometry and biology*. Cambridge, NY: Cambridge University Press; 1991.
9. Nasci RS. Relationship of wing length to adult dry weight in several mosquito species (Diptera: Culicidae). *J Med Entomol* 1990;**27**:716–9.
10. Siegel JP, Novak RJ, Lampman RL, Steinly BA. Statistical appraisal of the weight-wing length relationship of mosquitoes. *J Med Entomol* 1992;**29**(4):711–4.
11. Lehmann T, Dalton R, Kim E, Dahl E, Diabate A, Dabire R, et al. Genetic contribution to variation in larval development time, adult size, and longevity of starved adults of *Anopheles gambiae*. *Infect Genet Evol* 2006;**6**(5):410–6.

Downloaded From Vet-eBooks.com

12. Morales Vargas ER, Yaumphan P, Phumala-Morales N, Komalamisra N, Dujardin JP. Climate associated size and shape changes in *Aedes aegypti* (Diptera: Culicidae) populations from Thailand. *Infect Genet Evol* 2010;**10**(4):580−5.
13. Rohlf FJ. Rotational fit (Procrustes) methods [Special Publiation Number 2]. In: Rohlf F, Bookstein F, editors. *Proceedings of the Michigan morphometrics workshop*. Ann Arbor, MI: The University of Michigan Museum of Zoology; 1990. p. 227−36. p. 380.
14. Klingenberg CP. Multivariate allometry. In: *Advances in morphometrics. Proceedings of the 1993 NATO-ASI on morphometrics*. Marcus LF, Corti M, Loy A, Naylor GJP, Slice D, editors. NATO ASI, seR. A, Life sciences. New York: Plenum Publishers; 1996. p. 23−49.
15. Albrecht GH, Gelvin BR, Hartman SE. Ratios as a size adjustment in morphometrics. *Am J Phys Anthropol* 1993;**91**(4):441−68.
16. Burnaby TP. Growth-invariant discriminant functions and generalized distances. *Biometrics* 1966;**22**:96−110.
17. Dujardin JP, Slice DE. Contributions of morphometrics to medical entomology. In: Tibayrenc M, editor. *Encyclopedia of infectious diseases*. John Wiley & Sons, Inc.; 2006, ISBN 9780470114209. p. 435−47.
18. Zelditch ML, Swiderski DL, Sheets HD, Fink WL. *Geometric morphometrics for biologists: a primer*. New-York: Elsevier, Academic Press; 2004, ISBN 0-12-77846-08.
19. Adams DC, Rohlf FJ, Slice DE. Geometric morphometrics: ten years of progress following the "revolution". *Ital J Zool* 2004;**71**:5−16.
20. Rodhain F. Ecology of *Aedes aegypti* in Africa and Asia. *Bull Soc Pathol Exot* 1996;**89**(2):103−6.
21. Rohlf FJ, Bookstein FL. Computing the uniform component of shape variation. *Syst Biol* 2003;**52**(1):66−9.
22. Baylac M, Frieß M. Fourier descriptors, Procrustes superimposition, and data dimensionality: an example of cranial shape analysis in modern human populations. In: Slice DE, editor. *Modern morphometrics in physical anthropology*; 2005. p. 145−65.
23. Bookstein FL. Introduction to methods for landmark data [Special Publication No 2]. In: Rohlf FJ, Bookstein FL, editors. *Proceedings, Michigan morphometrics workshop, 1988*. Ann Arbor, MI: The University of Michigan Museum of Zoology; 1990. p. 216−25.
24. Bookstein FL. Landmark methods for forms without landmarks: localizing group differences in outline shape. *Med Image Anal* 1997;**1**:225−43.
25. Bookstein FL. Shape and the information in medical images: a decade of the morphometric synthesis. *Comput Vis Image Underst* 1997;**66**:97−118.
26. Gunz P, Mitteroeker P, Bookstein FL. Semilandmarks in three dimensions. In: Slice DE, editor. *Modern morphometrics in physical anthropology*. New York: Kluwer Academic/Plenum Publishers; 2005.
27. Yee WL, Sheets HD, Chapman PS. Analysis of surstylus and aculeus shape and size using geometric morphometrics to discriminate *Rhagoletis pomonella* and *Rhagoletis zephyria* (Diptera: Tephritidae). *Ann Entomol Soc Am* 2011;**104**(2):105−14. http://dx.doi.org/10.1603/AN10029.
28. Gunz P, Mitteroecker P. Semilandmarks: a method for quantifying curves and surfaces. *Hystrix* 2012. http://dx.doi.org/10.4404/hystrix-24.1-6292.
29. Rohlf FJ, Slice DE. Extensions of the Procrustes method for the optimal superimposition of landmarks. *Syst Zool* 1990;**39**:40−59.
30. Perez SY, Bernal V, Gonzalez PN. Differences between sliding semilandmark methods in geometric morphometrics, with an application to human craniofacial and dental variation. *J Anat* 2006;**208**:769−84.

Downloaded From Vet-eBooks.com

31. Dujardin JP, Kaba D, Solano P, Dupraz M, McCoy KD, Jaramillo ON. Outline-based morphometrics, an overlooked method in arthropod studies? *Infect Genet Evol* 2014;**28**: 704−14.
32. Boussès P, Dehecq JS, Brengues C, Fontenille D. Updated inventory of mosquitoes (Diptera: Culicidae) of the island of La Réunion, Indian Ocean. *Bull Soc Pathol Exot* 2013; **106**:113−25.
33. Francoy TM, Faria Franco F, Roubik DW. Integrated landmark and outline-based morphometric methods efficiently distinguish species of *Euglossa* (Hymenoptera, Apidae, Euglossini). *Apidologie* 2012;**43**:609−17.
34. Dujardin JP, Pham Thi K, Truong Xuan L, Panzera F, Pita S, Schofield CJ. Epidemiological status of kissing-bugs in South East Asia: a preliminary assessment. *Acta Trop* 2015;**151**:142−9.
35. Kuhl FP, Giardina CR. Elliptic fourier features of a closed contour. *Comput Graph Image Process* 1982;**18**:236−58.
36. Lohmann GP. Eigenshape analysis of microfossils: a general morphometric procedure for describing changes in shape. *Math Geol* 1983;**15**:659−72.
37. Rohlf FJ, Archie JW. A comparison of Fourier methods for the description of wing shape in mosquitoes (Diptera:Culicidae). *Syst Zool* 1984;**33**:302−17.
38. Lestrel PE. Methods for analyzing complex two-dimensional forms: elliptic Fourier functions. *Am J Hum Biol* 1989;**1**:149−64.
39. Lestrel PE. Introduction and overview of Fourier descriptors; Chapter 2. In: *Fourier descriptors and their applications in biology*. Cambridge: Cambridge University Press; 1997. p. 22−44.
40. Claude J. *Morphometrics with R*. Springer Science + Business Media, LLC; 2008, ISBN 978-0-387-77789-4.
41. Crampton JS. Elliptic Fourier shape analysis of fossil bivalves: some practical considerations. *Lethaia* 1995;**28**:179−86.
42. Firmat C, Gomes-Rodrigues H, Renaud S, Claude J, Hutterer R, Garcia-Talavera F, et al. Mandible morphology, dental microwear, and diet of the extinct giant rats *Canariomys* (Rodentia: Murinae) of the Canary Islands (Spain). *Biol J Linn Soc* 2010;**101**:28−40.
43. Huxley JS. *Problems of relative growth*. Methuen London; 1932.
44. Caro-Riaño H, Jaramillo N, Dujardin JP. Growth changes in *Rhodnius pallescens* under simulated domestic and sylvatic conditions. *Infect Genet Evol* 2009;**9**(2):162−8.
45. Arnqvist G, Mårtensson T. Measurement error in geometric morphometrics: empirical strategies to assess and reduce its impact on measure of shape. *Acta Zool Acad Sci Hung* 1998;**44**(1−2):73−96.
46. Klingenberg CP, McIntyre G. Geometric morphometrics of developmental instability: analyzing patterns of fluctuating asymmetry with Procrustes methods. *Evolution* 1998; **52**(5):1363−75.
47. Klingenberg CP, Leamy LJ. Quantitative genetics of geometric shape in the mouse mandible. *Evolution* 2001;**55**:2342−52.
48. Bitner-Mathé BC, Klaczko LB. Size and shape heritability in natural populations of *Drosophila mediopunctata*: temporal and microgeographical variation. *Genetica* 1999; **105**:35−42.
49. Orengo DJ, Prevosti A. Relationship between chromosomal polymorphism and wing size in a natural population of *Drosophila subobscura*. *Genetica* 2002;**115**:311−8.
50. Hatadani LM, Klaczko LB. Shape and size variation on the wing of *Drosophila mediopunctata*: influence of chromosome inversions and genotype-environment interaction. *Genetica* 2008;**133**:335−42.

Downloaded From Vet-eBooks.com

51. Breuker CJ, Patterson JS, Klingenberg CP. A single basis for developmental buffering of *Drosophila* wing shape. *PLoS One* 2006;**1**(1):e7.
52. Patterson J, Klingenberg C. Developmental buffering: how many genes? *Evol Dev* 2007; **9**(6):525−6.
53. Klingenberg CP, Leamy LJ, Cheverud JM. Integration and modularity of quantitative trait locus effects on geometric shape in the mouse mandible. *Genetics* 2004;**166**:1909−21.
54. Workman MS, Leamy LJ, Routman EJ, Cheverud JM. Analysis of quantitative trait locus effects on the size and shape of mandibular molars in mice. *Genetics* 2002;**160**(4): 1573−86.
55. Klingenberg CP, Leamy LJ, Routman EJ, Cheverud JM. Genetic architecture of mandible shape in mice: effects of quantitative trait loci analyzed by geometric morphometrics. *Genetics* 2001;**157**(2):785−802.
56. Iriarte PF, Norry FM, Hasson ER. Chromosomal inversions effect body size and shape in different breeding resources in *Drosophila buzzatii*. *Heredity* 2003;**91**:51−9.
57. Shrimpton A, Robertson A. The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*: I. Allocation of third chromosome bristle effects to chromosome sections. *Genetics* 1988;**118**:437−43.
58. Long A, Mullaney S, Reid L, Fry J, Langley C, Mackay TFC. High resolution mapping of genetic factors affecting abdominal bristle number in *Drosophila melanogaster*. *Genetics* 1995:139.
59. De la Riva J, Le Pont F, Ali V, Matias R, Mollinedo S, Dujardin JP. Wing geometry as a tool for studying the *Lutzomyia longipalpis* (Diptera: Psychodidae) complex. *Memórias do Inst O Cruz* 2001;**96**(8):1089−94.
60. Gumiel M, Catalá S, Noireau F, de Arias AR, Garcia A, Dujardin JP. Wing geometry in *Triatoma infestans* (Klug) and *T. melanosoma* Martinez, Olmedo and Carcavallo (Hemiptera: Reduviidae). *Syst Entomol* 2003;**28**(2):173−9.
61. Dujardin JP, Le Pont F, Baylac M. Geographic versus interspecific differentiation of sandflies: a landmark data analysis. *Bull Entomol Res* 2003;**93**:87−90.
62. Dujardin JP, Le Pont F. Geographic variation of metric properties within the neotropical sandflies. *Infect Genet Evol* 2004;**4**(4):353−9.
63. Camara M, Caro-Riaño H, Ravel S, Dujardin JP, Hervouet JP, de Meeus T, et al. Genetic and morphometric evidence for isolation of a tsetse (Diptera: Glossinidae) population (Loos islands, Guinea). *J Med Entomol* 2006;**43**(5):853−60.
64. Aytekin AM, Alten B, Caglar S, Ozbel Y, Kaynas S, Simsek FM, et al. Phenotypic variation among local populations of phlebotomine sand flies (Diptera: Psychodidae) in southern Turkey. *J Vector Ecol* 2007;**32**(2):226−34.
65. Henry A, Thongsripong P, Fonseca-Gonzalez I, Jaramillo-Ocampo N, Dujardin JP. Wing shape of dengue vectors from around the world. *Infect Genet Evol* 2010. http://dx.doi.org/10.1016/j.meegid.2009.12.001.
66. Jirakanjanakit N, Leemingsawat S, Dujardin JP. The geometry of the wing of *Aedes (Stegomyia) aegypti* in isofemale lines through successive generations. *Infect Genet Evol* 2008;**8**:414−21.
67. Jirakanjanakit N, Leemingsawat S, Thongrungkiat S, Apiwathnasorn C, Singhaniyom S, Bellec C, et al. Influence of larval density or food variation on the geometry of the wing of *Aedes (Stegomyia) aegypti*. *Trop Med Int Health* 2007;**12**(11):1354−60.
68. Daly HV. A statistical and empirical evaluation of some morphometric variables of honey bee classification. In: Foottit RG, Sorensen JT, editors. *Ordination in the study of morphology, evolution and systematics of insects: applications and quantitative genetic rationales*. New York: Elsevier; 1992. p. 127−56. p. 418.

Downloaded From Vet-eBooks.com

69. Anderson W. Genetic divergence in body size among experimental populations of *Drosophila pseudoobscura* kept at different temperatures. *Evolution* 1973;**2**(27): 278−84.
70. Partridge L, Barrie B, Fowler K, French V. Evolution and development of body size and cell size in *Drosophila melanogaster* in response to temperature. *Evolution* 1994;**48**: 1269−76.
71. Falconer DS. *Introduction to quantitative genetics*. London and New-York: Longman; 1981.
72. Houle D. Comparing evolvability and variability of quantitative traits. *Genetics* 1992; **130**(1):195−204.
73. Monteiro L, Diniz-Filho JA, Dos Reis SF, Araújo ED. Geometric estimates of heritability in biological shape. *Evolution* 2002;**56**(3):563−72.
74. Klingenberg C. Quantitative genetics of geometric shape: heritability and the pitfalls of the univariate approach. *Evolution* 2003;**57**(1):191−5.
75. Klingenberg C, Monteiro L. Distances and directions in multidimensional shape spaces: implications for morphometric applications. *Syst Biol* 2005;**54**(4):678−88.
76. Lande R. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution* 1979:402−16.
77. Mezey J, Houle D. The dimensionality of genetic variation for wing shape in *Drosophila melanogaster*. *Evolution* 2005:1027−38.
78. Hansen T, Houle D. Measuring and comparing evolvability and constraint in multivariate characters. *J Evol Biol* 2008;**21**(5):1201−19.
79. Klingenberg C. Morphological integration and developmental modularity. *Annu Rev Ecol Evol Syst* 2008:115−32.
80. Klingenberg C, Debat V, Roff DA. Quantitative genetics of shape in cricket wings: developmental integration in a functional structure. *Evolution* 2010;**64**(10):2935−51.
81. Marroig G. Morphological integration and modularity. *Annu Rev Ecol Evol Syst* 2016; **47**(1).
82. Roff DA, Mousseau TA. Quantitative genetics and fitness: lessons from *Drosophila*. *Heredity* 1987;**58**:103−18.
83. Gilchrist AS, Partridge L. The contrasting genetic architecture of wing size and shape in *Drosophila melanogaster*. *Heredity* 2001;**86**:144−52.
84. Hoffman AA, Shirriffs J. Geographic variation for wing shape in *Drosophila serrata*. *Evolution* 2002;**56**:1068−73.
85. Waddington CH. Genetic assimilation for an acquired character. *Evolution* 1953;**7**: 118−26.
86. Levit GS, Hossfeld U, Olssonza L. From the "Modern Synthesis" to cybernetics: Ivan Ivanovich Schmalhausen (1884−1963) and his research program for a synthesis of evolutionary and developmental biology. *J Exp Zool B Mol Dev Evol* 2006;**306B**:89−106.
87. Gibson G, Dworkin I. Uncovering cryptic genetic variation. *Nat Rev Genet* 2004;**5**: 681−90.
88. Bergman A, Siegal ML. Evolutionary capacitance as a general feature of complex gene networks. *Nature* 2003;**424**(6948):501−4.
89. Rutherford SL, Lindquist S. Hsp90 as a capacitor for morphological evolution. *Nature* 1998;**396**:336−42.
90. Debat V, Milton CC, Rutherford S, Klingenberg CP, Hoffmann AA. Hsp90 and the quantitative variation of wing shape in *Drosophila melanogaster*. *Evolution* 2006;**60**: 2529−38.

Downloaded From Vet-eBooks.com

91. Suzuki Y, Nijhout HF. Evolution of a polyphenism by genetic accommodation. *Science* 2006;**5761**(311):650−2.
92. Pennisi E. Evolution: hidden genetic variation yields caterpillar of a different color. *Science* 2006;**591a**.
93. Kitano H. Biological robustness. *Nat Rev Genet* 2004;**5**:826−37.
94. Wagner A. Neutralism and selectionism: a network-based reconciliation. *Nat Rev Genet* 2008;**9**(12):965−74.
95. Masel J. Q&A: evolutionary capacitance. *BMC Biol* 2013;**11**:103.
96. Usinger RL, Wygodzinsky P, Ryckman RE. The biosystematics of Triatominae. *Annu Rev Entomol* 1966;**11**:309−30.
97. Costa J, Peterson AT, Dujardin JP. Indirect evidences suggest homoploid hybridization as a possible mode of speciation in Triatominae (Hemiptera, Heteroptera, Reduviidae). *Infect Genet Evol* 2008;**9**(2):263−70.
98. Costa J, Bargues MD, Neiva VL, Lawrence GG, Gumiel M, Oliveira G, et al. Phenotypic variability confirmed by nuclear ribosomal DNA suggests a possible natural hybrid zone of *Triatoma brasiliensis* species complex. *Infect Genet Evol* 2016;**37**:77−87.
99. Costa J, Felix M. *Triatoma juazeirensis* sp. nov. from Bahia state, Northeastern Brazil (Hemiptera: Reduviidae: Triatominae). *Mem Inst Oswaldo Cruz* 2007;**102**:87−90.
100. Costa J, Monteiro F, Beard CB. *Triatoma brasiliensis* Neiva, 1911 the most important Chagas' disease vector in Brazil Phylogenetic and population analyzes correlated to epidemiologic importance. *Am J Trop Med Hyg* 2001;**65**:280.
101. Dujardin JP, Costa J, Bustamante D, Jaramillo N, Catalá S. Deciphering morphology in Triatominae: the evolutionary signals. *Acta Trop* 2009;**110**:101−11.
102. Kitthawee S, Dujardin JP. *Diachasmimorpha longicaudata*: reproductive isolation and geometric morphometrics of the wings. *Biol Control* 2009;**51**(1):191−7.
103. Dujardin JP, Beard CB, Ryckman R. The relevance of wing geometry in entomological surveillance of Triatominae, vectors of Chagas disease. *Infect Genet Evol* 2007;**7**(2): 161−7.
104. Dobzhansky T. *Genetics of evolutionary process*. New York: Colombia University Press; 1971. p. 505.
105. Schlichting CD, Pigliucci M. *Phenotypic evolution: a reaction norm perpective*. Sunderland, MA: Sinauer Associates, Inc.; 1998.
106. Hillesheim E, Stearns SC. The responses of *Drosophila melanogaster* to artificial selection on body weight and its phenotypic plasticity in two larval food environments. *Evolution* 1991;**45**(8):1909−23.
107. David JR, Moreteau B, Gauthier JP, Petavy G, Stockel A, Imasheva AG. Reaction norms of size characters in relation to growth temperature in *Drosophila melanogaster*: an isofemale lines analysis. *Gen Sel Evol* 1994;**26**:229−51.
108. Debat V, Begin M, Legout H, David JR. Allometric and nonallometric components of *Drosophila* wing shape respond differently to developmental temperature. *Evolution* 2003; **57**(12):2773−84.
109. Görür G. The importance of phenotypic plasticity in herbivorous insect speciation. In: Anantakrishnan TN, Whitman D, editors. *Insect phenotypic plasticity diversity of responses*. USA: Science Publishers, Inc. Enfield (NH); 2005. p. 145−72.
110. Dujardin JP, Panzera P, Schofield CJ. Triatominae as a model of morphological plasticity under ecological pressure. *Memórias do Inst Oswaldo Cruz* 1999;**94**:223−8.
111. Thompson JD. Phenotypic plasticity as a component of evolutionary change. *Trends Ecol Evol* 1971;**6**:246−9.

Downloaded From Vet-eBooks.com

112. Ananthakrishnan TN. Perspectives and dimensions of phenotypic plasticity in insects. In: Ananthakrishnan TN, Whitman D, editors. *Insect phenotypic plasticity diversity of responses*. USA: Science Publishers Inc. Enfield (NH); 2005. p. 1–23. p. 213.
113. Wilson EO, Brown WL. Revisionary notes on the *sanguinea* and *neogagates* groups of the ant genus *Formica*. *Psyche* 1955;**62**:108–29.
114. Brown WL, Wilson EO. Character displacement. *Syst Zool* 1956;**5**:49–64.
115. Grant PR. Convergent and divergent character displacement. *Biol J Linn Soc* 1972;**4**:39–68.
116. Losos JB. Ecological character displacement and the study of adaptation. *Proc Natl Acad Sci* 2000;**97**(11):5693–5.
117. Grant PR, Grant BR. Evolution of character displacement in Darwin's finches. *Science* 2006;**313**:224.
118. Adams DC, Rohlf FJ. Ecological character displacement in *Plethodon*: biomechanical differences found from a geometric morphometric study. *Proc Natl Acad Sci USA* 2000; **97**:4106–11.
119. Wilkins AS. Canalization: a molecular genetic perspective. *BioEssays* 1996;**19**:257–62.
120. Graham J, Freeman D, Emlen J. Developmental stability: a sensitive indicator of populations under 'stress'. In: Landis WG, Hugues JS, Lewis MA, editors. *Environmental toxicology and risk assessment, ASTM STP 1179*. Philadelphia: American Society for Testing and Materials; 1993. p. 136–58.
121. Debat V, David P. Mapping phenotypes: canalization, plasticity and developmental stability. *Trends Ecol Evol* 2001;**16**(10):555–61.
122. Debat V, Alibert P, David P, Paradis E, Auffray JC. Independence between developmental stability and canalization in the skull of the house mouse. *Proc R Soc Lond Ser B, Biol Sci* 2000;**267**(1442):423–30.
123. Réale D, Roff DA. Inbreeding, developmental stability, and canalization in the sand cricket *Gryllus firmus*. *Evolution* 2003;**57**(3):597–605.
124. Félix MA, Barkoulas M. Pervasive robustness in biological systems. *Nat Rev Genet* 2015; **16**(8):483–96.
125. Dworkin I. Canalization, cryptic variation, and developmental buffering: a critical examination and analytical perspective. In: Hallgrimsson B, Hall BK, editors. *Variation: a central concept in biology*. Oxford, UK: Academic Press; 2005. p. 131–58.
126. Palmer AR, Strobeck C. Fluctuating asymmetry: measurement, analysis, patterns. *Annu Rev Ecol Syst* 1986;**17**:391–421.
127. Nattero J, Dujardin JP, del Pilar FM, GR E. Host-feeding sources and habitats jointly affect wing developmental stability depending on sex in the major Chagas disease vector *Triatoma infestans*. *Infect Genet Evol* 2015;**36**:539–46.
128. Houle D. High enthusiasm and low r-squared. *Evolution* 1998;**52**(6):1872–6.
129. Leamy LJ, Klingenberg CP. The genetics and evolution of fluctuating asymmetry. *Annu Rev Ecol Evol Syst* 2005:1–21.
130. Van Dongen SV. Fluctuating asymmetry and developmental instability in evolutionary biology: past, present and future. *J Evol Biol* 2006;**19**(6):1727–43.
131. Hoffmann AA, Woods RE, Collins E, Wallin K, White A, McKenzie JA. Wing shape versus asymmetry as an indicator of changing environmental conditions in insects. *Aust J Entomol* 2005;**44**:233–43.
132. Mayr E. The biological species concept. In: Wheeler QD, Meier R, editors. *Species concepts and phylogenetic theory: a debate*. New York: Columbia University Press; 2000. p. 17–29.
133. Mayr E. *Principles of systematic zoology*. New York: McGraw-Hill; 1969.

Downloaded From Vet-eBooks.com

134. Hennig W. Phylogenetic systematics. *Annu Rev Entomol* 1965;**10**:97−116.
135. Wiley EO. The evolutionary species concept reconsidered. *Syst Zool* 1978;**27**:17−26.
136. Wiley EO, Mayden RL. The evolutionary species concepts. In: Wheeler QD, Meier R, editors. *Species concepts and phylogenetic theory: a debate*. New York: Columbia University Press; 2000. p. 70−92.
137. Mishler BD, Theriot EC. The phylogenetic species concept (sensu mishler and theriot): Monophyly, Apomorphy, and phylogenetic species concept. In: Wheeler QD, Meier R, editors. *Species concepts and phylogenetic theory: a debate*. New York: Columbia University Press; 2000. p. 44−54.
138. Baylac M, Villemant C, Simbolotti G. Combining geometric morphometrics with pattern recognition for the investigation of species complexes. *Biol J Linn Soc* 2003;**80**(1):89−98.
139. Becerra JM, Valdecasas AG. Landmark superimposition for taxonomic identification. *Biol J Linn Soc* 2004;**81**:267−74.
140. Dujardin JP. Morphometrics applied to medical entomology. *Infect Genet Evol* 2008;**8**: 875−90.
141. Perrard A, Baylac M, Carpenter JM, Villemant C. Evolution of wing shape in hornets: why is the wing venation efficient for species identification? *J Evol Biol* 2014;**27**:2665−75.
142. Villegas J, Feliciangeli MD, Dujardin JP. Wing shape divergence between *Rhodnius prolixus* from Cojedes (Venezuela) and *R. robustus* from Mérida (Venezuela). *Infect Genet Evol* 2002;**2**:121−8.
143. Matias A, De la Riva JX, Torrez M, Dujardin JP. *Rhodnius robustus* in Bolivia identified by its wings. *Memorias do Inst Oswaldo Cruz* 2001;**96**(7):947−50.
144. Ruangsittichai J, Apiwathnasorn C, Dujardin JP. Interspecific and sexual shape variation in the filariasis vectors *Mansonia dives* and *Ma. bonneae*. *Infect Genet Evol* 2011;**11**(8): 2089−94.
145. Jaramillo ON, Dujardin JP, Calle-Londoño D, Fonseca-González I. Geometric morphometrics for the taxonomy of 11 species of *Anopheles* (*Nyssorhynchus*) mosquitoes. *Med Vet Entomol* 2013;**29**(1):26−36.
146. Roggero A, Passerin d'Entrèves P. Geometric morphometric analysis of wings variation between two populations of the *Scythris obscurella* species-group: geographic or interspecific differences ? (Lepidotera: Scythrididae). *Shil Rev Lepidopterol* 2005;**33**(130):101−12.
147. Villemant C, Simbolotti G, Kenis M. Discrimination of Eubazus (Hymenoptera, Braconidae) sibling species using geometric morphometrics analysis of wing venation. *Syst Entomol* 2007;**32**(4):625−34.
148. Francuski L, Jasmina Ludõski J, Vujíc A, Milankov V. Wing geometric morphometric inferences on species delimitation and intraspecific divergent units in the *Merodon ruficornis* group (Diptera, Syrphidae) from the Balkan peninsula. *Zoolog Sci* 2009;**26**(4): 301−8.
149. Kitthawee S, Dujardin JP. The geometric approach to explore the Bactrocera tau complex (Diptera: Tephritidae) in Thailand. *Zoology* 2010;**113**(4):243−9.
150. Lyra ML, Hatadani LM, de Azeredo-Espin AM, Klaczko LB. Wing morphometry as a tool for correct identification of primary and secondary new world screwworm fly. *Bull Entomol Res* 2009;**23**:1−8.
151. Dujardin JP, Kaba D, Henry AB. The exchangeability of shape. *BMC Res Notes* 2010;**3**: 266. http://dx.doi.org/10.1186/1756−0500−3−266.
152. Jordaens K, Van Dongen S, Van Riel P, Geenen S, Verhagen R, Backeljau T. Multivariate morphometrics of soft body parts in terrestrial slugs: comparison between two datasets, error assessment and taxonomic implications. *Biol J Linn Soc* 2002;**75**(4):533−42.

Downloaded From Vet-eBooks.com

153. Rasmussen P, Wheeler W, Moser T, Vine L, Sullivan B, Rusch D. Measurements of Canada goose morphology: sources of error and effects on classification of subspecies. *J Wildl Manag* 2001;**65**(4):716−25.
154. Chivian E, Bernstein AS. Embedded in nature: human health and biodiversity. *Environ Health Perspect* 2004;**112**:A12−3.
155. Keesing F, Holt RD, Ostfeld RS. Effects of species diversity on disease risk. *Ecol Lett* 2006;**9**:485−98.
156. Shannon CE, Weaver W. *The mathematical theory of communication*. Urbana: University of Illinois Press; 1949.
157. Simpson EH. Measurement of diversity. *Nature* 1949;**163**:688.
158. Roy K, Balch DP, Hellberg ME. Spatial patterns of morphological diversity across the Indo-Pacific: analyses using strombid gastropods. *Proc R Soc Lond* 2001;**268**:2503−8.
159. Neige P. Spatial patterns of disparity and diversity of the recent cuttlefishes (Cephalopoda) across the Old World. *J Biogeogr* 2003;**30**:1125−37.
160. Foote M. Discordance and concordance between morphological and taxonomic diversity. *Paleobiology* 1993;**19**(2):185−204.
161. Dujardin JP, Thongsripong P, Henry AB. The mosquito Fauna: from metric disparity to species diversity. In: Wahl C, editor. *Morphometrics*. InTech; 2012, ISBN 978-953-51-0172-7.
162. Dujardin JP, Bermúdez H, Schofield CJ. The use of morphometrics in entomological surveillance of silvatic foci of Triatoma infestans in Bolivia. *Acta Trop* 1997;**66**:145−53.
163. Dujardin JP, Bermúdez H, Gianella A, Cardozo L, Ramos E, Saravia R, et al. Uso de marcadores genéticos en la vigilancia entomológica de la enfermedad de Chagas. In: Alfred Cassab J, Noireau F, Guillen G, editors. *La enfermedad de chagas en Bolivia conocimientos científicos al inicio del programa de control (1998−2002)*. La Paz: Ministerio de Salud y Prevision Social, OMS/OPS, IRD and IBBA; 1999. p. 157−69. p. 259.
164. Garros C, Dujardin JP. Genetic and phenetic approaches to Anopheles systematics. In: *Anopheles mosquitoes new insights into malaria vectors*; 2013. http://dx.doi.org/10.5772/56090.
165. Hernández ML, Dujardin JP, Gorla DE, Catalá SS. Potential sources of Triatoma infestans reinfesting peridomiciles identified by morphological characterization in los Llanos, La Rioja, Argentina. *Mem Inst Oswaldo Cruz* 2013:91−7.
166. Tapis M, Hausermann W. Demonstration of differential domesticity of *Aedes aegypti* (L.) (Diptera: Culicidae) in Africa by mark-releaserecapture. *Bull Entomol Res* 1975;**65**: 199−208.
167. Harrington L, Scott TW, Lerdthusnee K, Coleman RC, Costero A, Clark GG, et al. Dispersal of the dengue vector Aedes aegypti within and between rural communities. *Am J Trop Med Hyg* 2005;**72**(2):209−20.
168. Slatkin M. Rare alleles as indicators of gene flow. *Evolution* 1985;**39**(1):53−65.
169. Slatkin M. Estimating levels of gene flow in natural populations. *Genetics* 1981;**99**: 323−35.
170. McKay JK, Latta RG. Adaptive population divergence: markers, QTL and traits. *Trends Ecol Evol* 2002;**17**.
171. Spitze K. Population structure in Daphnia obtusa: quantitative genetic and allozymic variation. *Genetics* 1993;**135**:367−74.
172. Hiernaux J. Long-term biological effects of human migration from the African savanna to the equatorial forest: a case study of human adaptation to a hot and wet climate. In: Harrison GA, editor. *Population structure and human variation*. Cambridge: Cambridge University Press; 1977. p. 187−218.

Downloaded From Vet-eBooks.com

173. Rogers A, Harpending HC. Population structure and quantitative characters. *Genetics* 1983;**105**:985−1002.
174. Whitlock MC. Neutral additive variance in a metapopulation. *Genet Res* 1999;**74**:215−21.
175. de Paula AS, Diotaiuti L, Schofield CJ. Testing the sister-group relationship of the Rhodniini and Triatomini (Insecta: Hemiptera: Reduviidae: Triatominae). *Mol Phylogenet Evol* 2005;**35**:712−8.
176. Morales Vargas RE, Phumala-Morales N, Tsunoda T, Apiwathnasorn C, Dujardin JP. The phenetic structure of *Aedes albopictus*. *Infect Genet Evol* 2012;**13**:242−51.
177. Schachter-Broide J, Dujardin JP, Kitron U, Gürtler RE. Spatial structuring of Triatoma infestans (Hemiptera, Reduviidae) populations from northwestern Argentina using wing geometric morphometry. *J Med Entomol* 2004;**41**(4):643−9.
178. Gaspe MS, Gurevitz JM, Gürtler RE, Dujardin JP. Origins of house reinfestation with Triatoma infestans after insecticide spraying in the Argentine Chaco using wing geometric morphometry. *Infect Genet Evol* 2013;**17**:93−100.
179. Bouyer J, Ravel S, Dujardin JP, de Meeus T, Vial L, Thevenon S, et al. Population structuring of *Glossina palpalis gambiensis* (Diptera: Glossinidae) according to landscape fragmentation in the Mouhoun river, Burkina Faso. *J Med Entomology* 2007;**44**(5): 788−95.
180. Solano P, Kaba D, Ravel S, Dyer NA, Sall B, Vreysen MJB, et al. Population genetics as a tool to select tsetse control strategies: Suppression or eradication of *Glossina palpalis gambiensis* in the niayes of senegal. *PLoS Negl Trop Dis* 2010;**4**:e692.
181. Kaba D, Ravel S, Acapovi-Yao G, Solano P, Allou K, Bosson-Vanga H, et al. Phenetic and genetic structure of tsetse fly populations (*Glossina palpalis palpalis*) in southern Ivory Coast. *Parasites Vectors* 2012;**5**:153.
182. Feliciangeli MD, Sanchez-Martin M, Marrero R, Davies C, Dujardin JP. Morphometric evidence for a possible role of *Rhodnius prolixus* from palm trees in house re-infestation in the State of Barinas (Venezuela). *Acta Trop* 2007;**101**:169−77.
183. Fitzpatrick S, Feliciangeli M, Sanchez-Martin MJ, Monteiro FA, Miles MA. Molecular genetics reveal that silvatic *Rhodnius prolixus* do colonise rural houses. *PLoS Negl Trop Dis* 2008;**2**(4):e210.

Downloaded From Vet-eBooks.com

This page intentionally left blank

Downloaded From Vet-eBooks.com

Evolution of Resistance to Insecticide in Disease Vectors

14

P. Labbé[1], *J.-P. David*[2], *H. Alout*[1,3], *P. Milesi*[1], *L. Djogbénou*[4], *N. Pasteur*[1], *M. Weill*[1]

[1]Institut des Sciences de l'Evolution de Montpellier (UMR 5554, CNRS-UM-IRD-EPHE), Université de Montpellier, Montpellier, France; [2]Laboratoire d'Ecologie Alpine (UMR 5553 CNRS-UGA), Université Grenoble—Alpes, Grenoble, France; [3]Colorado State University, Fort Collins, CO, United States; [4]Université d'Abomey Calavi, Cotonou, Benin

1. Introduction

The control of vector-borne diseases represents one of the greatest global public health challenges of the 21st century. They contribute substantially to the global burden of infectious diseases (~17%) and their prevalence tends to increase (World Health Organization[1]). Human population growth in many areas has led to extensive deforestation, irrigation, and urbanization, and these environmental modifications have created conditions that favor the proliferation of many arthropod vectors, such as mosquitoes, ticks, flies, and so on. Primarily in developing countries, 3.2 billion people are now at risk for contracting many new or reemerging diseases.[2]

Mosquitoes are probably the most common vectors of infectious diseases (review in Ref. 3); 3500 species are found throughout the World and, in almost all species, the females find the proteins they need for developing eggs through blood-feeding on vertebrates. This makes mosquitoes particularly prone to transfer viruses and other parasites between humans and animals hosts. They are vectors of malaria and arboviruses (dengue, yellow fever, zika, Japanese encephalitis, west nile, and chikungunya). Other major vector-borne diseases (sleeping sickness, leishmaniasis, onchocerciasis, plague, Bartonellosis, rickettsioses, Lyme disease, ehrlichiosis, babesiosis, anaplasmosis, trypanosomiasis, Chagas disease, and several viral diseases) are transmitted by non-mosquito arthropods (tsetse flies (*Glossina* sp.), sand flies (Phlebotominae), black flies (Simulidae), houseflies, fleas, lice, cockroaches, and Triatomine bugs).

Some tropical vector-borne diseases have been observed in developed countries (e.g., Chikungunya or West Nile virus in Europe and USA). If climate (temperature, rainfall, and humidity) does influence disease transmission, expansion of disease range is mostly due to human factors, such as forest clearing, increased travel, transport, and economical activities (e.g., the geographic distribution of *Aedes albopictus* has considerably increased through worldwide commerce of used tires and because of its capacity of diapausing and the resistance of its eggs to desiccation[4]). Overall, it seems that the main determinants of vector-borne diseases' prevalence are socioeconomic (see Refs. 5—7). Unfortunately, the burden that vector-borne diseases impose directly impairs the

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00014-7
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

public health and socioeconomic development of many of the poor areas. Controlling these diseases is thus a necessity. This ideally entails active case-detection and treatment of human infections (vaccines, antiparasitic drugs). However, few vaccines are currently available (e.g., for yellow fever, Japanese encephalitis) and many pathogens, such as *Plasmodium*, are now resistant to antiparasitic drugs. Moreover, populations from endemic countries struggle to get access to them, notably due to economic impediments. Thus in many instances, the control of vectors is the only affordable measure.

The first documented attempts to control malaria by limiting the densities of vectors go back to the Roman times: in an attempt to control the "Roman fever" (the name of malaria at that time), Julius Caesar himself had the Codetan swamp around Rome drained and planted with trees (Varro about 40 BC[8]). While such environmental modifications aiming at reducing the number of breeding sites have shown great success, today the most common and affordable way of fighting the major disease vectors is the use of insecticides.[9–11] Many scientific investigations and reports show that the use of synthetic insecticides can dramatically reduce the risk of insect-borne diseases. Insecticides, combined with extensive use of drugs, have rapidly led to the eradication of many diseases (e.g., malaria) from most nontropical areas of the world, but in spite of initial successes, eradication has proven more elusive in the tropics.[12] However, mechanisms allowing survival to insecticide exposures have been selected in many species of arthropod vectors. Resistance to all classes of synthetic insecticides is now widespread among pests of public health importance, and it is considered to be the most important impediment in the successful control of vector-borne diseases.

2. Insecticide Resistance: Definition and History

Insecticide resistance in pest populations affects both economy and public health at a worldwide scale: it decreases crop yields (and thus profitability), induces the need to increase the quantity of insecticide and to develop new insecticides (thereby having a strong impact on costs and on the environment), and finally it is responsible for higher incidence of human or animal diseases.[13,14] This general society problem, however, provides evolutionary biologists with a unique contemporary model, ideal for studying how new adaptations evolve by natural selection. The selecting agent is known (insecticides), evolution is recent and rapid (few years after insecticide selection), and the biological and genetic mechanisms are often known (see Part 3). This explains why it has been the subject of such a large body of work over the years.

Resistance is defined as a heritable decrease of the susceptibility to an insecticide.[15] Three categories of resistance can be distinguished: behavioral (avoidance of contact with insecticide), physiological (e.g., increased cuticle thickness), and biochemical (enhanced insecticide detoxification and sequestration and/or decreased insecticide target sensitivity). Few examples of behavioral (e.g., *Anopheles gambiae* on Bioko Island and Senegal[16,17] or *Anopheles funestus* in Benin and Tanzania[18–20]) and physiological resistances have been reported; whether they are heritable remains debated, and it is difficult to assess the level of protection they provide. Biochemical resistances

Downloaded From Vet-eBooks.com

typically result in relatively high level of protection and are genetically determined. Resistant individuals carry one or several genetic mutations that prevent insecticide disruption of the target functioning. As a result, the frequency of resistance gene(s)/allele(s) increases in the population over time. Insecticide resistance is confirmed by toxicological tests (bioassays) establishing resistance ratio (or RR corresponding to the number by which an insecticide dose must be multiplied in order to obtain the same mortality in resistant than in susceptible insects). It can be investigated at many levels, from the molecular characterization of genes/alleles conferring resistance and their biochemical products, to the effect of these genes on the fitness (i.e., mean reproductive success) of the individuals carrying resistance alleles, to the dynamics and evolution of these resistance alleles in natural vector populations and their effect on disease control.

The first recorded attempt of insect pest control, the application of tobacco juice against sheep scabs, is found in the literature of the 18th century.[21] The first case of resistance was reported in 1908, in a population of San Jose scale (*Aspidiotus perniciosus*) resistant to lime sulfur.[22] A century later (2007), 553 arthropod species were reported as resistant to at least one insecticide, among many disease vectors. More than 100 mosquito species are resistant to at least one insecticide (including 56 *Anopheline* species, 39 *Culicine* species); *Culex pipiens pipiens* and *Anopheles albimanus* are resistant to more than 30 different compounds.[14]

2.1 Synthetic Insecticides

Originally, only inorganic insecticides (such as lime sulfur) and natural products were available, for example, flower-extracted pyrethrum for malaria control in the 1930s. Today, four classes of organic (synthetic) insecticides are essentially used: the organochlorines (OCs), the organophosphates (OPs), the carbamates (CXs), and the pyrethroids (PYRs), with, respectively, 4429, 1375, 30, and 414 metric tonnes of active ingredient used annually for global vector control from 2000 to 2009.[23]

The first synthetic insecticides, introduced during World War II for malaria control, belonged to the OC class. The first one was the dichlorodiphenyltrichloroethane or DDT (introduced in 1943), which targets the voltage-gated sodium channels (Na-channels); another was the cyclodiene (CD) dieldrin, which targets the γ-aminobutyric acid (GABA) receptor; both targets being essential in the insect nervous system (see Part 3). In addition to their public health applications, enormous tonnages of DDT and dieldrin were used worldwide in agriculture. It was at first a great success with large WHO-led campaigns leading to reduction of morbidity and mortality from malaria in many endemic regions after World War II. Widely acclaimed, DDT and dieldrin rapidly selected resistance in insect vectors. In *An. gambiae*, resistance to DDT was first noted 11 years after its introduction,[24] while a population from northern Nigeria was reported resistant to dieldrin soon after.[25] DDT resistance has now been reported in mosquitoes (*Aedes* sp., *Anopheles* sp., and *Culex* sp.), houseflies, sand flies, body lice, and head lice, while resistance to dieldrin (60% of reported cases of resistance before 1990) has been detected in more than 277 arthropods, including mosquitoes (*Aedes* sp., *Anopheles* sp., and *Culex* sp.), fleas, ticks, biting flies, bedbugs, cockroaches, and human lice.[1,9,10,26,27]

Downloaded From Vet-eBooks.com

An important issue against these insecticides was their environmental impact. Rachel Carson's book "Silent Spring"[28] was a seminal work publicizing and politicizing the toxic effects of the accumulation of DDT and its metabolites in the food chain. In vertebrates, DDT can interfere with reproduction, and in humans it can have neurologic, carcinogenic, and reproductive effects, although the evidences remain debated. These insecticides are also extremely stable in the environment, contaminating groundwater and remaining in soil long after their use. In the 1970s, the Persistent Organic Pollution Treaty led to total banning of dieldrin and to the banning of DDT for all uses except malaria control when this disease is very frequent and there is no alternative. DDT use rapidly declined in the 1970s (it is no longer used in Latin America),[29] but it gained new advocates due to the development of resistance to the alternative insecticides, and to its low cost.[1,29–32] Consequently, its use quadrupled between 2007 and 2009.[23]

From the late 1970s, OCs were replaced by the PYRs class of vector control, and these became widely used in agriculture and public health, and more particularly against malaria vector. They are today by far the most-used insecticides, with 81% of the World spray coverage.[23] As DDT, these insecticides target the Na-channels (i.e., neurotoxic effect). Their rapid popularity comes from their very low toxicity to human, their rapid knock-down (KD) effect associated with an excitorepellancy effect. PYR-based indoor residual spraying (IRS) and insecticide-treated nets and curtains (ITNs) are currently advocated as standard malaria vector control strategies.[1]

PYR resistance was reported in 1993, in *An. gambiae* populations from Côte d'Ivoire[33] and later in *C. pipiens quinquefasciatus* also in West Africa.[34] Resistance is now widespread in mosquitoes (*Aedes* sp., *Anopheles* sp., and *Culex* sp. (see Ref. 35 for a review)), body and head lice, ticks (e.g., *Boophilus microplus*), and fleas.[1,9,10] As PYR resistance developed, many control programs attempted to revert to DDT for disease control. However, these insecticides share a common target site, and there is cross-resistance to both insecticide classes in many locations.[30,32]

Finally, two other classes of synthetic insecticides are used at a large scale worldwide: the OPs and the CXs, which were first used in the 1940s and the 1950s, respectively.[1,15] OPs and CXs target the synaptic acetylcholinesterase (AChE), an essential enzyme in the nervous system. They are usually used as larvicids (although some are now considered for ITN impregnation and IRS as an alternative to PYR[36]), and are particularly well suited for species with delimited breeding sites. However, they have a short half-life, and two to three rounds of IRS are needed per year. This, combined in some instances with their high price, can make these insecticides too costly for most malaria control programs, despite fewer reports of resistance.[32] Early resistance to these insecticides has been detected shortly after their first application: for example, first OP treatments in the Montpellier area (southern France) started in 1969, the first resistance being detected only 3 years later.[37] Resistance has now been recorded in mosquitoes (*Aedes* sp., *Anopheles* sp., and *Culex* sp.), biting flies (e.g., *Simulium damnosum*, vector of onchocerciasis), sand flies, houseflies, and fleas (reviews in Refs. 1,10,26,27).

During 2006–2008, few new insecticides were described: neonicotinoids, phthalic acid diamides, or anthranilic acid diamides; however, they are used mostly for agricultural pests, not for disease vectors.[14,38] Finally, another type of synthetic

Downloaded From Vet-eBooks.com

insecticides is growth regulators (GR). It regroups synthetic products called juvenoids that mimic the juvenile hormone (JH) (review in Ref. 39) and chitine inhibitors (see Ref. 40). So far, only few cases of resistance have been reported in houseflies and mosquitoes (e.g., resistance to methoprene, a JH analog in the mosquito *Ochlerotatus nigromaculi*).[41]

In summary, most often only PYRs are available, essentially for economic cost reasons: the most recent PYR had been introduced in mid-1980s and no new synthetic insecticide has been found since mid-1990s. The shrinking availability of insecticides as a result of resistance is exacerbated by the removal from the market of insecticides that are no longer registered for public health use: some compounds are too costly, and insecticide use is restricted by regulatory agencies, due to environmental concerns. Consequently, new environment-proof products (high selectivity, no effects on nontargets) are now required for sustainable vector control.[42]

2.2 Alternative Insecticides

Environmental pollution concerns and unresolved issues pertaining to the toxicity of synthetic insecticides to humans and nontarget species have led the public and researchers' interest to investigate alternative "biological" insecticides.[43] Three main types of these alternative insecticides are documented: (1) bacterial toxins, (2) essential oils, and (3) fungi.

There are two main sources of bacterial toxins: *Bacillus sphaericus* (Bs) and *Bacillus thuringiensis* (Bt). They kill insect larvae by producing proteic toxins binding to various receptors on midgut epithelial cells (review in Ref. 39). Bs toxicity is due to a binary toxin, whereas Bt toxicity is due to the interaction of four different toxins. These larvicides are presented as highly specific and effective at low doses, and are thus expected to be safe for the environment. Toxins extracted from Bs and a variety of Bt (*B. thuringiensis* var *israelensis* or Bti) are used for mosquito control. In these species, bacterial toxins show some differences in specificity: Bti is more effective against *Aedes* and *Culex* species than against *Anopheles*, whereas Bs is more effective against *Culex* than *Anopheles* species, and has no effect on *Aedes* species that lack receptors. While the presence of several toxins was expected to delay resistance apparition, Bs and Bti resistances have been detected in various mosquitoes,[43–45] and resistance to Bt has been detected in several agricultural pests.[46]

Although less documented, essential oils are investigated as potential biological larvicides. They are advocated to be more specific than synthetic insecticides, and biodegradable, thus with reduced impact on the environment. Variable efficacies seem to represent a restraint for pest control; identifying the bioactive components instead of raw products could be the solution to this problem (for review see Ref. 47).

Finally, fungi can be used as biological insecticides: they target the adult stage of mosquitoes and are used essentially for malaria control. The fungus *Metarhizium anisopliae* has been shown to reduce *An. gambiae* adult life span in the laboratory and in the field in Tanzania,[48] while *Beauveria bassiana* decreases the survival of

Downloaded From Vet-eBooks.com

another malaria vector, *Anopheles stephensi*.[49] These agents have several advantages: they are cheap, easily stored for long term, and specific to insects. These fungal insecticides have a direct effect on *Plasmodium* transmission and are expected to decrease malaria prevalence. Finally, their acting late in life is considered by several authors to be an important advantage, as it will decrease selective pressure and reduce the risk of resistance development (potentially "evolution-proof" insecticides[42,50,51]).

To conclude this part, it should be noted that insecticide resistance does not appear in all treated species, at least on the short term. This can be linked to the particular life cycle of the species or to molecular constraints preventing the evolution of resistance mechanism. For example, after decades of treatment, the tsetse flies (*Glossina* sp.) have not yet developed resistance to DDT or PYRs, probably due to their very small number of youngs, which limits their evolutionary reactivity.[1,10,52] Similarly, for several years, *Aedes aegypti* did not develop the most efficient resistance mechanism to OPs and CXs (i.e., insensitiveAChE) because its particular codon usage prevented the apparition of the required mutation[53]; the presence of the mutation was, however, described in India in 2015.[54] This last example shows that understanding why resistance occurs or not also requires elucidating the mechanisms of insecticide resistance at the molecular and biochemical levels.

3. Mechanisms of Resistance

The targets of most insecticides are critical proteins of the insect nervous system. Insecticides bind to specific sites on their targets and disrupt their function. Any mechanism that decreases the insecticide effect will lead to resistance. This encompasses reduced penetration of the insecticide, increased excretion or sequestration of the insecticide, increased metabolism of the insecticide, and finally target modification that limits the binding of the insecticide. However, a behavioral change resulting in a reduced exposure to the insecticide can also be viewed as a resistance mechanism, if it is heritable: for example, *Anopheles* mosquitoes have been reported to have changed their blood-feedings habits, by seeking hosts outdoor (exophily and exophagy) rather than indoor (endophily and endophagy[16–20]); however, whether this behavior is heritable remains debated.

The first three mechanisms are poorly documented and do not seem to play a prominent role in resistance.[55] Most studies aiming at understanding the mechanisms and the genetic bases of insecticide resistance focus on metabolic resistance and target-site modification. Usually, these resistances are explained by a limited number of mechanisms, monogenic in the case of insecticide target modifications.

In this chapter, we present the various documented mechanisms of resistance. We specifically focus on disease vector species, although many mechanisms are common to agricultural pests. We insist on the evolutionary aspects of resistance, while the detailed mechanisms are treated more succinctly, and only for the major ones. More comprehensive reviews can be found (e.g., Refs. 27,35,39,55–57). Moreover, the recent explosion of genomic studies on resistance frustrates any pretention to exhaustiveness.

Downloaded From Vet-eBooks.com

3.1 Metabolic Resistance

Metabolic resistance regroups the various mechanisms that lead to the degradation of the insecticide in less- or nontoxic products, thus decreasing the quantity of toxic molecules that reach the target. These so-called "detoxification enzymes" belong mainly to three large gene families, cytochrome P450 monooxygenases (P450s or CYPs for genes), glutathione S-transferases (GSTs), and carboxylesterases (COEs), and most studies focus on a small set of genes. Genomic studies can, however, access mechanisms that had previously proven intractable. They allow deeper description of known resistance gene families and help find new candidate genes. They have suggested that other enzyme families may be implicated, such as UDP-glycosyl-transferases (UGTs), sulfotransferases, aldehyde dehydrogenases, NADH-cytochrome *b* reductases, NADH dehydrogenases, NADH-ubiquinone oxidoreductases, nitrilase thioredoxin peroxidases, and cuticular genes (e.g., Refs. 56,58−60). However, in most cases the causal role of the candidates remains to be formally validated.

Detoxification enzymes are frequently divided into phase I and phase II enzymes depending on their role in detoxification pathways with hydrolases and oxidases acting during phase I, and transferases acting during phase II.[60] These enzymes can act individually, synergically, or sequentially through complex insecticide degradation pathways. Such complexity is accentuated by the redundancy of insect detoxification systems. A given detoxification enzyme may indeed metabolize different insecticides (although with different kinetic parameters), thus contributing to cross-resistance. On the other hand, different enzymes may degrade the same insecticide, and contribute in an additive manner to the resistance phenotype. In natural populations, several metabolic mechanisms can be present in the same species (e.g., Ref. 61), and metabolic resistance often combines with target-site modifications leading to high-resistance levels and complex cross-resistance patterns.

3.1.1 Glutathione S-Transferases

Various xenobiotics contain the tripeptide glutathione; GSTs catalyze the reaction of the sulfhydryl group of this tripeptide. This sulfhydryl group reacts with electrophilic sites on xenobiotics, leading to formation of conjugates that are more readily excreted and typically less toxic than the parent insecticide. In addition to this direct detoxification, GSTs play a role in phase II detoxification (see later).

GST enzymes are present in most insects. They represent a large family of generalist detoxifying enzymes (six classes of GSTs have been identified in the genome of *An. gambiae*) and have thus broad substrate specificities. The GST family expands either by alternative splicing or by local gene duplication, the last leading to clusters of GST genes.

GSTs are primarily associated with resistance to OCs, particularly DDT, and OPs. GST-based resistance seems to be associated with an increased amount of enzyme resulting from gene duplication or, more often, upregulation. A constitutive GST overexpression was frequently reported in mosquito populations showing elevated resistance level to DDT.[55,61−63] Quantitative genetic analyses identified a quantitative

Downloaded From Vet-eBooks.com

trait locus (QTL) for resistance to DDT in *An. gambiae*, within which there is a cluster of eight GSTs.[64] Among them, GSTE2 was then shown to metabolize DDT.[63] GSTE2 ortholog in *Ae. aegypti* and *An. funestus* was further shown to metabolize DDT.[65,66]

GSTs are also suspected to play a role in the resistance to PYRs in mosquitoes through sequestration. Lumjuan et al.[67] showed that the partial KD of *Ae. aegypti* GSTE2 and GSTE7 led to an increased susceptibility to the PYR deltamethrin. Similarly, Riveron et al.[66] show that GSTE2 contributes to PYR resistance in *An. funestus* probably through sequestration.

3.1.2 Cytochrome P450 Monooxygenases

Cytochrome P450 monooxygenases (P450) are heme-thiolate enzymes found in all living organisms.[68] They are best known for their monooxygenase activity, but they can catalyze a wide range of reactions. In insects, P450s are associated with the metabolism of endogenous compounds, such as hormones, and are involved in the phase I detoxification of a variety of xenobiotics including plant toxins, pollutants, and chemical insecticides.[57,69,70] P450s are frequently represented by more than a 100 genes in insect genomes, so that the identification of those involved in insecticide resistance is challenging. Some of them are inducible by xenobiotics and expressed at higher level in classical detoxification tissues (midgut, fat bodies, Malpighian tubules), although such properties do not ensure their actual contribution to insecticide resistance. Insecticide resistance is often linked to the overexpression of one or multiple P450s through upregulation or gene amplification, although mutations may also lead to resistance.

P450s have been reported as responsible for resistance to most insecticide classes, particularly DDT, PYRs, and CXs. In addition, some P450s are also capable of activating particular OPs, such as malathion and diazinon (i.e., they become toxic when oxidized). The contribution of P450s in insecticide resistance can be estimated by combining the exposure of insects to the P450 inhibitor piperonyl butoxide (PBO) and subsequent bioassays with insecticides: if P450s are implicated, the resistance level is usually decreased in the presence of PBO. However, PBO does not equally inhibit all P450s, so that absence of PBO-induced resistance decrease does not mean that no P450 is implicated. The role of P450s in resistance may also be evidenced by biochemical assays measuring either the global heme content,[10] or more specific enzyme activities using known P450 substrates, such as ethoxycoumarin (ECOD method) or resorufin (EROD method). However, biochemical assays are not always capable of detecting P450-based resistance, because these assays have a low specificity, unlike some P450s.

Following the sequencing of mosquito genomes and the development of microarrays,[61] transcriptomics has been intensively used for detecting overtranscribed P450s in resistant populations, leading to the identification of several CYP genes associated with resistance in mosquitoes and other insects (reviews in Refs. 27,55,57,71,72). In mosquitoes, some of them were validated as capable of contributing to insecticide metabolism by functional approaches, such as heterologous expression followed by in vitro insecticide metabolism, RNA interference, or transgenic expression. These include the *Anopheles* genes CYP6Z1, CYP6M2, CYP6P3,

Downloaded From Vet-eBooks.com

CYP6P9, CYP6P4, CYP6P7, CYP6AA3[73–78]; *Aedes* genes CYP9J32, CYP9J24, CYP9J28, CYP6BB2[79–81]; and the *Culex* gene CYP9M10.[82] Interestingly, it was shown that *Anopheles* CYP6M2 and CYP6P3 can metabolize insecticides from different classes, supporting the role of P450s in cross-resistance, and raising concerns for insecticide-resistance management.[74,83] Although gene expression studies have identified multiple P450s overexpressed in resistant insects, very few data are available on the genetic factors controlling their overexpression. Recently, the use of targeted deep sequencing allowed the identification of gene amplifications controlling the overproduction of P450s in multiple PYR-resistant population of *Ae. aegypti* worldwide.[60] High-throughput sequencing approaches also allowed identifying nonsynonymous variations affecting P450s potentially linked to insecticide detoxification.[60]

3.1.3 Carboxylesterases

More than 30 genes coding COEs are found in insects (see detailed review in Refs. 26,39). Most COEs are serine esterases, that is, they have a serine residue within a catalytic triad necessary for hydrolysis. COEs bind to an ester group and then break the ester bound by a process of acylation−deacylation. Multiple forms of COEs are found in insects, with broad and overlapping substrate specificities.

The majority of insecticides, including almost all CXs and OPs, most PYRs, and some GRs bear ester groups. In most cases, hydrolysis of the ester group leads to a reduced toxicity of the insecticide. Consequently, COEs are often involved in metabolic resistance mechanisms, although the level of resistance conferred is relatively low (∼10×) compared to target-site resistance. As for P450s, the role of COEs in resistance is usually diagnosed by the addition of a synergist, the S,S,S-tributyl phosphorotrithioate (DEF) to bioassays. DEF inhibits COEs (but also GSTs): if COEs contribute to resistance, insecticide toxicity is expected to increase in the presence of DEF, significantly more in resistant than in susceptible insects.[84] COE-based resistance has been detected in various species, mainly against OPs and to a lesser extent to PYRs.[10,15]

OP resistance in *Culex* mosquitoes is generally caused by an elevated COE protein quantity, up to 80 times the level found in susceptible individuals.[85] Two esterases, α-esterase (or esterase A) and β-esterase (or esterase B), have been recognized based on their higher affinity for, respectively, α- and β-naphthylacetate.[86] Their overexpression is usually caused by an increased gene copy number (gene amplification) of one or both esterases, although upregulation may also contribute to overexpression.[87,88] The loci coding for the esterases A and B behave as a single locus named *Ester*.[89] The number of gene copies within an amplification of the *Ester* locus can vary greatly, potentially in relation with the intensity of insecticide treatments.[84,90,91]

Amplified esterases have also been described in the mosquitoes *An. gambiae* and *Ae. aegypti* in association with resistance to the OP temephos.[92,92a] Orthologs of these genes were also found amplified in association with temephos resistance in the tiger mosquito *Ae. albopictus*.[93] Biochemical assays also pointed out the role of esterases in PYR hydrolysis in mosquitoes,[94] although no particular esterase has yet been validated as able to hydrolyze PYRs. Moreover, it appears that the PYR metabolites

Downloaded From Vet-eBooks.com

produced by esterases could be further metabolized by overexpressed P450s of the subfamily CYP6Z in PYR-resistant populations, suggesting synergy between these two resistance mechanisms.[77]

Because overexpressed COEs can represent a large percentage of the total protein of the insect (up to 12% of the soluble proteins in some resistant mosquitoes[95]), it is difficult to disentangle their sequestration effect (i.e., binding to the insecticide without hydrolysis) from the direct hydrolysis of the insecticide. This appears to depend on the species and the esterase allele: hydrolysis appears predominant in the aphid E4 esterase, while in mosquitoes the $Ester^{B1}$ and $Ester^{2}$ alleles rather sequester the insecticide and show a lower hydrolysis activity.[96–98] However, qualitative changes affecting COEs may also be responsible for resistance to particular insecticides. For example, resistance to the OP malathion in Anophelinae, *Musca domestica* and *Lucilia cuprina*, was associated with particular point mutations inducing a faster hydrolysis.[10,26,56,99]

In terms of population genetics, COE resistance to OPs in *C. pipiens* is probably one of the best-studied cases. In this species, resistance to OPs was monitored since late 1960s in the Montpellier area of Southern France.[100–102] This long-term monitoring showed that several *Ester*-resistance alleles have been replacing each other across time: $Ester^{1}$ was the first detected resistance allele in 1972, then $Ester^{4}$ in 1986, and finally $Ester^{2}$ arrived by migration in 1991. These alleles were selected in insecticide-treated areas, but also showed a fitness disadvantage or cost in absence of insecticide (lower mating success, lower survival, and so on).[103–108] The quantification of their fitness cost showed that the various alleles correspond to different fitness tradeoffs: $Ester^{4}$ was first favored over $Ester^{1}$ because of a lower cost (selection for a generalist allele). Then $Ester^{2}$ appeared to be replacing the first two alleles because it confers a higher resistance level, despite its relatively high cost (selection for a specialist allele[102]). Overall, this example confirms that insecticide resistance is a dynamic process, as new haplotypes can be selected for adjusting the resistance phenotype and the fitness of resistant individuals to insecticide pressures and environmental factors.

3.2 Target-Site Modification

Resistance by target-site modification is due to point mutations in the insecticide target gene that results in reduced binding of insecticides, rather than to a change in expression level. Because most insecticide targets are vital molecules, there is generally only a limited number of mutations in the target able to decrease insecticide affinity without impeding its original function to an unsustainable degree (see detailed review in Ref. 39). A mutation conferring resistance while partly impairing the target's normal function leads to a fitness cost.

3.2.1 GABA Receptors

GABA is a major neurotransmitter in the insect's central and peripheral nervous system and in neuromuscular junctions. The GABA receptors are linked to chlorine-gated channels, causing hyperpolarization that blocks the nervous influx. GABA receptors

Downloaded From Vet-eBooks.com

are the target of CDs. CDs are noncompetitive inhibitors that bind to a site on the receptor close to the chlorine-gated channel, stabilizing it in an inactive closed state. This induces an overexcitation by removal of the inhibition, and leads to convulsions and death of the insect. GABA receptors have also secondary-binding sites for some PYRs or insecticides of the avermectin family.[10]

Resistance to CDs is due to a decreased sensitivity to insecticide of the GABA receptor A, through a point mutation causing an amino acid change in the receptor-coding gene. This gene, called *Rdl* (Resistance to dieldrin, the most-used CD), has been first cloned in *Drosophila melanogaster*. In all *D. melanogaster*—resistant individuals, the *Rdl* locus displays a similar mutation at position 302 in the channel-lining domain sequence, changing an alanine into a serine (A302S). The role of this mutation in CD resistance was confirmed by directed mutagenesis. The serine residue occupies the insecticide-binding site of the GABA receptor and destabilizes its conformation (review in Ref. 109). The resistance allele (Rdl^R) is semidominant and can confer cross-resistance to other insecticides, such as fipronil (e.g., Refs. 56,109).

Due to an extensive use of CDs before their banning in the 1980s, resistance has been selected in several insect species, which all display a mutation at the same position (A302S or A302G).[56,109] Whether these mutations are costly depends on species: a fitness cost associated with resistance has been identified in *L. cuprina*[110] and has been suggested in *C. pipiens* and *An. albopictus*,[111,112] but no cost has been found in *D. melanogaster*,[109] even if resistance affects temperature sensitivity. The *Rdl* locus has been found duplicated in the greenbug *Myzus persicae*[113] and in a strain of *D. melanogaster*.[114] In the latter, a tandem duplication of 113 kb associates a susceptible and a resistance copy of the locus. The phenotype associated to this duplication was shown to be close to that of a standard heterozygote, namely an intermediate resistance level and a reduced heat shock recovery time.[114]

3.2.2 Voltage-Gated Sodium Channels

Nerve action potentials are transmitted by a wave of depolarization along the neural axon. They are due to the movement of sodium ions (Na^+) crossing the axonal membrane through the opening of voltage-gated sodium channels (VGSCs), and stop when these channels are inactivated. VGSCs are glycoproteins with a pore for ion transport and can adopt three different states: resting, open, or inactivated; the Na^+ ions pass only when the channels are open.[115]

VGSC are the targets of DDT and PYRs. When these insecticides bind to the VGSC, they slow their closing speed, prolonging the depolarization.[115–117] The intensity of the effect is dose-dependent, proportional to the number of Na-channels inactivated.[115] For PYRs, the magnitude of the effect depends on the type of insecticide molecules, type I (e.g., permethrin) or type II (e.g., lambda-cyhalothrin and deltamethrin), which, respectively, lack or not a cyano group. During action potential, type II PYRs lengthen the sodium flux more than type I, and thus usually display a more intense effect.[116] At the phenotypic level, inactivation of VSGC results in a rapid KD effect, the insect being incapacitated for some time, followed by recovery or death,

Downloaded From Vet-eBooks.com

depending on the species and development stages (in mosquitoes, the adults tend to recover, while larvae will drown).

One major mechanism, named knockdown resistance (kdr), is responsible for PYR and DDT resistance, by reducing the receptors sensitivity (binding capacity) to these insecticides and modifying the action potential of the channel.[39,117,118] First discovered in *M. domestica*, this mechanism has been described in many agricultural pests and vectors. This resistance mechanism has several consequences: it decreases the irritant and the repellent effects, and either cancels or reduces the KD effect.[119]

Extension mutations affecting the VGSC gene are called *kdr* mutations. By sequencing the VGSC protein (>2000 amino acids), the first two *kdr* mutations were identified in *M. domestica*, both in the second protein domain. The first one (L1014F) is associated with moderate (10−30×) PYR resistance; the second (M918T, also called *super-kdr*) is always associated with the L1014F and confers a higher resistance (up to 500×).[120] Substitution of the L1014 is found in a large variety of species (L1014F or L1014S, and also L1014H in *Heliothis virescens*) and corresponds to the kdr^R alleles.[116,117,121,122]

The phenotype conferred by kdr^R is recessive or semirecessive,[10,119] with higher resistance to type I than type II PYRs.[123] However, the various mutations show some specificity, as L1014F confers a high resistance to both DDT and permethrin (PYR), while L1014S confers a lower resistance to permethrin than to DDT.[121,124] Other *kdr* mutations (about 30 in total) have been described in various species, including *super-kdr* mutations.[116,117] Some of these mutations are conserved over a large array of organisms, while others are more specific and unique. In *Ae. aegypti*, the *kdr* phenotype has been observed, but it appears that a codon bias prevents the appearance of any L1014 mutation.[125] However, several other mutations have been observed associated with resistance in *Ae. aegypti* (e.g., V1023G/I, I1018 M/V, F1565C, D1794Y, or S996P[126]). In *Ae. albopictus*, the F1565C mutation has been observed, while no mutation has been found at the 1018 site.[81] The importance of these various mutations in the different resistance phenotypes is thus still in debate.

The role of the L1014 F/S mutations (kdr^R) as the sole cause of the *kdr* phenotype is still discussed.[127] kdr^R is clearly associated to PYR and DDT resistance in *Blatella germanica*, *C. pipiens*, houseflies, hornflies, and some moths (review in Ref. 128). In *An. gambiae*, although metabolic resistance is often present, high resistance to PYR and DDT is most of the times associated with a high kdr^R frequency, and resistant insects carry at least one kdr^R copy.[124,129−132] Moreover, kdr^R frequency usually increases when PYRs are used[133−135]: two alleles are spreading in *An. gambiae* in Africa, L1014F and L1014S mutations and analyses of the noncoding regions of the *kdr* gene suggest that the two alleles occurred several times independently (at least three times for L1014F and two times for L1014S[122,136,137]). Similarly, in West African *C. p. quinquefasciatus*, resistance frequency follows a gradient of treatment intensity.[34]

In the field, *An. gambiae* resistance to PYRs through *kdr* can lead to reduced repellent effect and decreased mortality. For example, kdr^R frequency is high in Benin and Côte d'Ivoire, while no other PYR-resistance mechanism was found (although they could have been overlooked): studies have shown strong diminution of vector control

Downloaded From Vet-eBooks.com

with PYR-treated bed nets in these countries.[138] In contrast, other studies have found that despite the high correlation between *kdr* mutations and PYR resistance, PYR-treated bed nets remained somewhat efficient against resistant *An. gambiae*.[127,139] This could be due to the ability of resistant mosquitoes to stay on a treated bed net longer than susceptibles, and thus absorb a high-enough quantity of insecticide to be killed.[119] For example, in Kenya, the use of PYR-treated bed nets increased kdr^R frequency, but had no impact on malaria and mosquito population densities, as both decreased in treated and untreated villages.[133] Similarly two studies found that kdr^R alone (i.e., in the absence of metabolic resistance) did not reduce bed net efficiency against resistant *An. stephensi*, despite a reduced KD effect.[140] The issue of the impact of *kdr* resistance on PYR-treated bed net efficiency to control malaria thus remains hotly debated.

3.2.3 Acetylcholinesterase

In the cholinergic synapses of invertebrate and vertebrate central nervous system, AChE terminates the synaptic transmission by rapidly hydrolyzing the neurotransmitter acetylcholine (ACh). AChE is the target of OPs and CXs insecticides, which are competitive inhibitors of ACh: when they bind to AChE, their very slow release prevents hydrolysis of the natural substrate. Consequently, ACh remains active in the synaptic cleft and the nervous influx is continued, leading to insect death by tetany.

In most insects there are two genes, *ace-1* and *ace-2*, coding for AChE1 and AChE2, respectively. In these species, AChE1 is the main synaptic enzyme while the physiological role of AChE2 is still uncertain. Diptera of the Cyclorrapha group or "true" flies (such as *D. melanogaster* and *M. domestica*) possess a single AChE, which is encoded by the *ace-2* gene and is the synaptic enzyme in that case. Phylogenetic analyses have shown that the presence of two *ace* genes is probably the ancestral insect state.[141,142]

The first molecular studies on an insensitive AChE conferring resistance to OPs and CXs were carried out on *D. melanogaster*. Several mutations were identified, each giving a low resistance when alone, and a higher resistance when in combination.[143] Similar results were later found with other Diptera that have only the *ace-2* gene (e.g., *M. domestica*[26]).

In mosquitoes where AChE1 is the synaptic enzyme, the most common resistance mutation (G119S) in the *ace-1* gene is located just near the active site. In *C. pipiens*, G119S occurred at least 3 times independently, once in *C. p. pipiens* and twice in *C. p. quinquefasciatus*.[53,144,145] However, two other mutations in *ace-1* have been identified, both close to the active site: (1) F331W has been observed only in *Culex tritaeniorhynchus*,[146,147] (2) F290V has been observed only in *C. p. pipiens*.[148,149] The type of mutation appears highly constrained by the codon use: until recently the G119S mutation was never found in *Ae. aegypti*, *Ae. albopictus*, or *C. tritaeniorhynchus*, probably because it requires two mutational steps.[53] It was, however, described in *Ae. aegypti* from India in 2015, apparently through a mutation from a different codon (R119S[54]).

Downloaded From Vet-eBooks.com

The *ace* mutations are responsible for a decreased inhibition of the AChE by the insecticides.[150] There are only few resistance mutations observed in various species, suggesting high constraints: those observed in the field are within the active gorge of the enzyme and cause steric problems with bulkier side-chains, while other substitutions (lab-engineered) often result in the inability of enzyme to degrade ACh.[26] The G119S *ace-1* mutation has recently been shown to interact synergistically with an unknown sex-linked gene to allow a >40 000-fold resistance to chlorpyrifos (OP[151]). Similarly, the G119S mutation associated with the kdr^R allele confers higher-resistance levels in *An. gambiae* to both OPs and CXs insecticides.[152]

The evolution of insensitive AChE1 has been studied in depth in the mosquitoes *C. pipiens* and *An. gambiae*. In *C. pipiens*, it was first detected in Southern France in 1978, 9 years after the beginning of OP treatments.[153] The gene coding for this G119S mutated AChE1 ($ace\text{-}1^R$) rapidly spread in treated natural populations. However, its frequency remained low in adjacent untreated areas connected by migration, indicating a fitness cost associated with $ace\text{-}1^R$.[104] The >60% reduction of AChE1 activity in G119S-resistant mosquitoes[154] may probably explain, at least partially, this cost, which is expressed phenotypically through various developmental and behavioral problems in individuals carrying $ace\text{-}1^R$.[105,107,108] Similarly, the F290V mutation is probably associated with a fitness cost, although it does not appear to be due to activity reduction.[149] Several independent heterogeneous duplications of the *ace-1* gene, putting a susceptible and a resistant copy in tandem ($ace\text{-}1^D$), have been identified in *C. p. pipiens* and *C. p. quinquefasciatus*.[145,155] These alleles are thought to be selected because they confer an alternative fitness trade-off, that is, reducing the cost of the $ace\text{-}1^R$ allele, but with a decreased resistance level as well.[156] However, some $ace\text{-}1^D$ can be associated to extremely deleterious phenotypes when homozygotes.[156,157] Several other duplications have been observed recently in the Mediterranean area, with a F290V copy instead of a G119S copy.[149] In *An. gambiae*, the occurrence of $ace\text{-}1^R$ has been detected in several West African countries, and this allele is probably spreading from a single origin.[144,158,159] As in *C. pipiens*, this mutation is associated with a strong selective cost in *An. gambiae*.[160] A duplication carrying a G119S copy has also been found, and appears to follow the same trajectory as in *C. pipiens*[161]; the *An. gambiae* $ace\text{-}1^D$ allele also provides an alternative phenotype, a reduced cost associated with a reduced resistance.[160] In both species, it has been suggested that the relative fitness of the two alleles ($ace\text{-}1^R$ and $ace\text{-}1^D$) may depend on the intensity of insecticide treatments.[156,160] Finally, two studies reported in 2015 have suggested that resistance alleles with multiple $ace\text{-}1^R$ copies are segregating in Africa[162,163]; the fitness consequences of such duplications remain, however, unknown.

3.3 Other Resistance Mechanisms

3.3.1 Growth Regulators

Juvenoids mimic JH and disrupt insect development. Few resistance cases have been described in various species (review in Ref. 39). High resistance to methoprene has

Downloaded From Vet-eBooks.com

been described in the mosquito *Ochlerotatus nigromaculis* in California, potentially through target-site mutation,[41] while a 7.7-fold resistance to the same insecticide has been reported in *C. p. pipiens* from New York.[45]

3.3.2 Toxin Receptors

Bt toxins have a complex mode of action not clearly understood. Bt resistance is increasing in the field in several pests.[46] Presently, the only report of field resistance in mosquito is a 33-fold resistance to Bti (Bt var. *israelensis*, the only Bt variety active on mosquitoes) detected in a natural population of *C. p. pipiens* from New York. However, the mechanism of this resistance was not investigated.[45] Genomic studies suggested several candidates for Bti resistance in *Ae. aegypti*, but they are not yet validated.[164] Finally, it appears that depending on the environmental conditions, some of the four Bti toxins may be inactivated,[165] which could favor the emergence of full Bti resistance through intermediate bouts of selection to each toxin independently.

For Bs toxins, resistance has been described essentially in mosquitoes of the *C. pipiens* complex, due to mutation in the toxin receptor. It developed very rapidly within the first year of treatment in India (10−155× resistance[43]) and in Tunisia (*Sp-T* gene, >5000× resistance[166]). Similarly, control using Bs toxins started in the early 1990s in Southern France and first failure was reported in 1994 in Port-Louis (near Marseille). This resistance (>10,000×) was due to a recessive sex-linked gene, named *sp-1*. In 1996, Bs resistance was reported close to the Spain border (Perpignan, 200 km away from Port St Louis); it was due to a second gene, *sp-2*, which was recessive and sex-linked.[167] Now Bs resistance has been observed worldwide in the *C. pipiens* complex.[43] Two of the alleles identified ($sp\text{-}2^R$ and an allele selected in a laboratory strain from California[168]) change the toxin receptor binding properties, and were found to be due to "stop" mutations or mobile element insertion in the toxin receptor.[169,170] The effect of the other alleles is unknown.[166] Bs resistance has also been selected in the laboratory in *An. stephensi*.[43]

3.4 Resistance Generalities

Some general patterns can be identified from the variety of mechanisms observed for insecticide resistance.

A first characteristic is that resistance evolves rapidly, with fast selective sweeps in field populations. Most of the times, resistance alleles are present in the field before insecticide treatments, at very low frequencies. They are selected locally but can spread very rapidly. A single resistance gene may have a large distribution,[71,109,122] for example, the worldwide migration of $Ester^2$ in *C. pipiens*.[101] Alternatively, other resistance alleles have multiple origins: $ace\text{-}1^R$ mutations in *C. pipiens* (G119S[144] or F290V[149]) or *kdr* mutations in *Ae. aegypti*.[171−174]

It also seems that resistance evolution is quite constrained. For target-site resistance, most mutations are costly and compromise the performance of the native protein function, so that codon usage may prevent resistance apparition.[53,125]

Downloaded From Vet-eBooks.com

Another issue is the cross-resistance. Cross-resistances between insecticide classes can be associated with the sharing of target sites. For example, kdr^R causes cross-resistance between DDT and PYRs in *An. gambiae*,[123] and $ace\text{-}1^R$ between OPs and CXs.[150] Cross-resistance can even be a greater issue when considering metabolic resistance. First, different genes belonging to a same enzyme family can cause resistance to several insecticides ("gene family cross-resistance"), even from different classes: for example, different COE and P450 genes cause resistance to DDT, others to PYRs, OPs, and CXs in *Anophelines*.[175] However, a unique gene may also be involved in resistance to several insecticides, from different classes: this is the case, for example, of the CYP6M2 gene (P450), which can metabolize both deltamethrin (PYR) and DDT (OC[57]). The consequences of these cross-resistances are a severe reduction of the availability of alternative insecticides, thereby gravely endangering vector control.

Finally, despite advances, a full analysis of resistance remains challenging due to the complexity of interactions, pleiotropy, and redundancy when several resistance mechanisms and/or resistance genes are present in the same insect.[39] Interactions between resistance loci have been studied in houseflies or mosquitoes, and most of them appear to be synergistic. Such synergies have been observed, for example, in *C. pipiens* between COE and *ace-1* for OP resistance,[176] between *ace-1* and an unknown gene, raising resistance to chlorpyrifos by more than 2000-fold compared to *ace-1* alone (>40,000-fold compared to susceptible[151]) and between *kdr* and P450 for PYR resistance,[82] in *Ae. aegypti* between repellents (DEET) and CXs,[177] in *An. gambiae* s.s. between *ace-1* and *kdr* for OPs and CXs resistance[152] or in three *Anopheles* species between PYR resistance and susceptibility to fungus applications.[42] Moreover, these interactions may vary with environmental conditions (positive synergism for resistance in treated area but negative synergism for cost in nontreated areas) or with the genetic background of the insect.[82] For example, the presence of kdr^R decreases the cost of $ace\text{-}1^R$ in *C. pipiens*.[178]

4. Conclusion

The natural history of mosquito-borne diseases is complex, and the interplay of climate, ecology, vector biology, and many other factors defies simplistic analyses. The recent resurgence of many of these diseases is a major cause for concern. Its principal determinants are politics, economics, and human activities (rather than climate change). In order to control these diseases and ameliorate the socioeconomic burden they cause in developing countries, vector control remains a powerful and accessible tool. However, any disease control strategy should take into account insecticide-resistance management as it can greatly impact its success (vector control failures) and may have a direct effect on pathogen transmission.[179–182] This includes first establishing a continuous survey of resistance at a local scale by implicating the local population, a difficult but essential task to set goals and evaluate success. Several survey sites in different conditions are required for sentinel

Downloaded From Vet-eBooks.com

purposes, together with some baseline information, to rapidly detect resistance, identify the mechanisms, and change the policies adequately.[183] In order to achieve this survey, basic tools, such as bioassays, remain most powerful, and should always be a preliminary step before more complex and more costly analyses. However, specific and validated molecular markers for the known resistance alleles (e.g., *kdr*, *ace-1*, and some metabolic markers) are also required to rapidly identify the origin and follow the dynamics of resistance at a minimum cost. These local surveys should then be integrated at a more global scale for vector control coordination, allowing informed decisions for using alternative tools to insecticides and preserving the remaining insecticides by carefully planning their use to minimize resistance selection. Clearly, the greatest challenge for successful vector and disease control is the coordination of the different actors (chemical industries, researchers, politics, control agencies, and local populations), which do not have the same agendas, motivations, or economical interests.

Besides its implications in public health and development, insecticide resistance remains a powerful evolutionary biology model to study the contemporary adaptation of organisms to a changing environment. It indeed allows a complete and integrative study, from the molecular mechanisms to the fitness consequences at the individual level and their impacts on insect population dynamics and interactions with pathogens. Moreover, it is for once pleasant to see that these rather fundamental approaches of evolutionary biology may have a direct impact in the society and help design new strategies for the successful control of some of the most threatening human diseases.[50]

References

1. WHO. *Pesticides and their applications for the control of vectors and pests of public health importance*. 6th ed. World Health Organization; 2006.
2. WHO. *World health statistics*. 2015.
3. Tolle MA. Mosquito-borne diseases. *Curr Probl Pediatr Adolesc Health Care* 2009;**39**: 97−140.
4. Enserink M. Entomology: a mosquito goes global. *Science* 2008;**320**:864−6.
5. Kay B, Vu SN. New strategy against *Aedes aegypti* in Vietnam. *Lancet* 2005;**365**:613−7.
6. Ooi E-E, Goh K-T, Gubler DJ. Dengue prevention and 35 years of vector control in Singapore. *Emerg Infect Dis* 2006;**12**:887−93.
7. Morrison AC, Zielinski-Gutierrez E, Scott TW, Rosenberg R. Defining challenges and proposing solutions for control of the virus vector *Aedes aegypti*. *PLoS Med* 2008;**5**:e68.
8. Cheesman DF. Varro and the small beasts: a bimillennium for microbiologists. *Nature* 1964;**203**:911−2.
9. Roberts DR, Andre RG. Insecticide resistance issues in vector-borne disease control. *Am J Trop Med Hyg* 1994;**50**:21−34.
10. Hemingway J, Ranson H. Insecticide resistance in insect vectors of human disease. *Annu Rev Entomol* 2000;**45**:371−91.
11. Beier J, Keating J, Githure JI, Macdonald M, Impoinvil D, Novak R. Integrated vector management for malaria control. *Malar J* 2008;**7**:S4.

Downloaded From Vet-eBooks.com

12. Dialynas E, Topalis P, Vontas J, Louis C. Miro and IRbase: IT tools for the epidemiological monitoring of insecticide resistance in mosquito disease vectors. *PLoS Negl Trop Dis* 2009;**3**:e465.
13. Georghiou GP, Lagunes-Tejeda A. *The occurrence of resistance to pesticides in arthropods*. Rome: Food and Agriculture Organization; 1991.
14. Whalon ME, Mota-Sanchez D, Hollingworth RM. *Global pesticide resistance in arthropods*. Cambridge, MA: CABI Publishing; 2008.
15. Nauen R. Insecticide resistance in disease vectors of public health importance. *Pest Manag Sci* 2007;**63**:628–33.
16. Reddy MR, Overgaard HJ, Abaga S, Reddy VP, Caccone A, Kiszewski AE, et al. Outdoor host seeking behaviour of *Anopheles gambiae* mosquitoes following initiation of malaria vector control on Bioko Island, Equatorial Guinea. *Malar J* 2011;**10**:184.
17. Ndiath MO, Mazenot C, Sokhna C, Trape J-F. How the malaria vector *Anopheles gambiae* adapts to the use of insecticide-treated nets by African populations. *PLoS One* 2014;**9**: e97700.
18. Moiroux N, Gomez MB, Pennetier C, Elanga E, Djènontin A, Chandre F, et al. Changes in *Anopheles funestus* biting behavior following universal coverage of long-lasting insecticidal nets in benin. *J Infect Dis* 2012;**206**:1622–9.
19. Russell TL, Govella NJ, Azizi S, Drakeley CJ, Kachur SP, Killeen GF. Increased proportions of outdoor feeding among residual malaria vector populations following increased use of insecticide-treated nets in rural Tanzania. *Malar J* 2011;**10**:80.
20. Sougoufara S, Diédhiou SM, Doucouré S, Diagne N, Sembène PM, Harry M, et al. Biting by *Anopheles funestus* in broad daylight after use of long-lasting insecticidal nets: a new challenge to malaria elimination. *Malar J* 2014;**13**:125.
21. Wood RJ. In: Bishop JA, Cook LM, editors. *Insecticide resistance: genes and mechanisms*. London: Academic Press; 1981. p. 53–96.
22. Melander A. Can insects become resistants to sprays? *J Econ Entomol* 1914;**7**:167–72.
23. WHO. *Global insecticide use for vector-borne disease control, a 10 year assessment (2000–2009)*. 2011. ISBN:9789241502153.
24. WHO. Malaria section. *Bull World Heal Organ* 1957;**16**:874.
25. Davidson G. Insecticide resistance in *Anopheles gambiae* Giles: a case of simple Mendelian inheritance. *Nature* 1956;**178**:863–4.
26. Oakeshott JG, Devonshire AL, Claudianos C, Sutherland TD, Horne I, Campbell PM, et al. Comparing the organophosphorus and carbamate insecticide resistance mutations in cholin- and carboxyl-esterases. *Chem Biol Interact* 2005;**157–158**:269–75.
27. Feyereisen R, Dermauw W, Van Leeuwen T. Genotype to phenotype, the molecular and physiological dimensions of resistance in arthropods. *Pestic Biochem Physiol* 2015;**121**: 61–77.
28. Carson R. *Silent spring*. Boston: Houghton Mifflin; 1962.
29. van den Berg H. Global status of DDT and its alternatives for use in vector control to prevent disease. *Environ Health Perspect* 2009;**117**:1656–63.
30. Brooke BD, Kloke G, Hunt RH, Koekemoer LL, Temu EA, Taylor ME, et al. Bioassay and biochemical analyses of insecticide resistance in southern African *Anopheles funestus* (Diptera: Culicidae). *Bull Entomol Res* 2001;**91**:265–72.
31. Rogan WJ, Chen A. Health risks and benefits of bis(4-chlorophenyl)-1,1,1-trichloroethane (DDT). *Lancet* 2005;**366**:763–73.
32. Coleman M, Casimiro SLR, Hemingway J, Sharp B. Operational impact of DDT reintroduction for malaria control on *Anopheles arabiensis* in Mozambique. *J Med Entomol* 2008;**45**:885–90.

Downloaded From Vet-eBooks.com

33. Elissa N, Mouchet J, Rivière F, Meunier JY, Yao K. Resistance of *Anopheles gambiae s.s.* to pyrethroids in Cote d'ivoire. *Ann Soc Belg Med Trop* 1993;**73**: 291–4.
34. Chandre F, Darriet F, Darder M, Cuany A, Doannio JMC, Pasteur N, et al. Pyrethroid resistance in *Culex quinquefasciatus* from West Africa. *Med Vet Entomol* 1998;**12**: 359–66.
35. Liu N, Xu Q, Zhu F, Zhang L, Nannan LIU, Qiang XU, et al. Pyrethroid resistance in mosquitoes. *Insect Sci* 2006;**13**:159–66.
36. Oxborough RM, Mosha FW, Matowo J, Mndeme R, Feston E, Hemingway J, et al. Mosquitoes and bednets: testing the spatial positioning of insecticide on nets and the rationale behind combination insecticide treatments. *Ann Trop Med Parasitol* 2008;**102**: 717–27.
37. Pasteur N, Sinègre G. Esterase polymorphism and sensitivity to Dursban organophosphorous insecticide in *Culex pipiens pipiens* populations. *Biochem Genet* 1975;**13**: 789–803.
38. Nauen R. Insecticide mode of action: return of the ryanodine receptor. *Pest Manag Sci* 2006;**62**:690–2.
39. Hollingworth RM, Dong K. In: Whalon ME, Mota-Sanchez D, Hollingworth RM, editors. *The biochemical and molecular genetic basis of resistance in arthropods*. Cambridge, MA: CAB International; 2008. p 192.
40. Hirose T, Sunazuka T, Omura S. Recent development of two chitinase inhibitors, Argifin and Argadin, produced by soil microorganisms. *Proc Jpn Acad Ser B Phys Biol Sci* 2010; **86**:85–102.
41. Cornel AJ, Stanich MA, McAbee RD, Mulligan FS. High level methoprene resistance in the mosquito *Ochlerotatus nigromaculis* (Ludlow) in Central California. *Pest Manag Sci* 2002;**58**:791–8.
42. Farenhorst M, Mouatcho JC, Kikankie CK, Brooke BD, Hunt RH, Thomas MB, et al. Fungal infection counters insecticide resistance in African malaria mosquitoes. *Proc Natl Acad Sci* 2009;**106**:17443–7.
43. Mittal PK. Biolarvicides in vector control: challenges and prospects. *J Vector Borne Dis* 2003;**40**:20–32.
44. Nielsen-Leroux C, Pasquier F, Charles JF, Sinegre G, Gaven B, Pasteur N. Resistance to *Bacillus sphaericus* involves different mechanisms in *Culex pipiens* (Diptera: Culicidae) larvae. *J Med Entomol* 1997;**34**:321–7.
45. Paul A, Harrington LC, Zhang L, Scott JG. Insecticide resistance in *Culex pipiens* from New York. *J Am Mosq Control Assoc* 2005;**21**:305–9.
46. Tabashnik BE, Brévault T, Carrière Y. Insect resistance to Bt crops: lessons from the first billion acres. *Nat Biotechnol* 2013;**31**:510–21.
47. George DR, Finn RD, Graham KM, Sparagano OA. Present and future potential of plant-derived products to control arthropods of veterinary and medical significance. *Parasit Vectors* 2014;**7**:28.
48. Scholte E-J, Ng'habi K, Kihonda J, Takken W, Paaijmans K, Abdulla S, et al. An Entomopathogenic fungus for control of adult African malaria mosquitoes. *Science* 2005;**308**: 1641–2.
49. Blanford S, Chan BHK, Jenkins N, Sim D, Turner RJ, Read AF, et al. Fungal pathogen reduces potential for malaria transmission. *Science* 2005;**308**:1638–41.
50. Michalakis Y, Renaud F. Malaria: evolution in vector control. *Nature* 2009;**462**:298–300.
51. Read AF, Lynch PA, Thomas MB. How to make evolution-proof insecticides for malaria control. *PLoS Biol* 2009;**7**:e58.

Downloaded From Vet-eBooks.com

52. Welburn SC, Coleman PG, Maudlin I, Fevre EM, Odiit M, Eisler MC. Crisis, what crisis? Control of Rhodesian sleeping sickness. *Trends Parasitol* 2006;**22**:123−8.
53. Weill M, Berticat C, Lutfalla G, Pasteur N, Philips A, Fort P, et al. Insecticide resistance: a silent base prediction. *Curr Biol* 2004;**14**:R552−3.
54. Muthusamy R, Shivakumar MS. Susceptibility status of *Aedes aegypti* (L.) (Diptera: Culicidae) to temephos from three districts of Tamil Nadu, India. *J Vector Borne Dis* 2015; **52**:159−65.
55. Nkya TE, Akhouayri I, Kisinza W, David J-P. Impact of environment on mosquito response to pyrethroid insecticides: facts, evidences and prospects. *Insect Biochem Mol Biol* 2013;**43**:407−4013.
56. Hemingway J, Hawkes NJ, McCarroll L, Ranson H. The molecular basis of insecticide resistance in mosquitoes. *Insect Biochem Mol Biol* 2004;**34**:653−65.
57. David J, Ismail HM, Chandor-proust A, Paine MJI. Role of cytochrome P450s in insecticide resistance : impact on the control of mosquito-borne diseases and use of insecticides on Earth. *Philos Trans R Soc B Biol Sci* 2013;**368**:20120429.
58. Oakeshott JG, Home I, Sutherland T, Russell R. The genomics of insecticide resistance. *Genome Biol* 2003;**4**:202.
59. Li X, Schuler MA, Berenbaum MR. Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annu Rev Entomol* 2007;**52**:231−53.
60. Faucon F, Dusfour I, Gaude T, Navratil V, Boyer F, Chandre F, et al. Identifying genomic changes associated with insecticide resistance in the dengue mosquito *Aedes aegypti* by deep targeted sequencing. *Genome Res* 2015;**25**:1347−59.
61. Strode C, Wondji CS, David J-P, Hawkes NJ, Lumjuan N, Nelson DR, et al. Genomic analysis of detoxification genes in the mosquito *Aedes aegypti*. *Insect Biochem Mol Biol* 2008;**38**:113−23.
62. Ranson H, Hemingway J. Mosquito glutathione transferases. In: *Gluthione transferases and gamma-glutamyl transpeptidases* vol. 401; 2005. p. 226−41.
63. Enayati AA, Ranson H, Hemingway J. Insect glutathione transferases and insecticide resistance. *Insect Mol Biol* 2005;**14**:3−8.
64. Ranson H, Claudianos C, Ortelli F, Abgrall C, Hemingway J, Sharakhova MV, et al. Evolution of supergene families associated with insecticide resistance. *Science* 2002;**298**: 179−81.
65. Lumjuan N, McCarroll L, Prapanthadara L, Hemingway J, Ranson H. Elevated activity of an Epsilon class glutathione transferase confers DDT resistance in the dengue vector, *Aedes aegypti*. *Insect Biochem Mol Biol* 2005;**35**:861−71.
66. Riveron JM, Yunta C, Ibrahim SS, Djouaka R, Irving H, Menze BD, et al. A single mutation in the GSTe2 gene allows tracking of metabolically based insecticide resistance in a major malaria vector. *Genome Biol* 2014;**15**:R27.
67. Lumjuan N, Rajatileka S, Changsom D, Wicheer J, Leelapat P, Prapanthadara L, et al. The role of the *Aedes aegypti* Epsilon glutathione transferases in conferring resistance to DDT and pyrethroid insecticides. *Insect Biochem Mol Biol* 2011;**41**:203−9.
68. Werck-Reichhart D, Feyereisen R. Cytochromes P450: a success story. *Genome Biol* 2000;**1**. Reviews 3003.
69. Feyereisen R. In: Gilbert LI, Iatrou K, Gill SS, editors. *Insect cytochrome P450*. Oxford, UK: Elsevier; 2005. p. 1−77.
70. Despres L, David J-PP, Gallet C, Després L, David J-PP, Gallet C. The evolutionary ecology of insect resistance to plant chemicals. *Trends Ecol Evol* 2007;**22**:298−307.
71. Ffrench-Constant RH, Daborn PJ, Le Goff G. The genetics and genomics of insecticide resistance. *Trends Genet* 2004;**20**:163−70.

Downloaded From Vet-eBooks.com

72. Vontas J, Kioulos E, Pavlidi N, Morou E, della Torre A, Ranson H. Insecticide resistance in the major dengue vectors *Aedes albopictus* and *Aedes aegypti*. *Pestic Biochem Physiol* 2012;**104**:126−31.
73. Chiu T-L, Wen Z, Rupasinghe SG, Schuler MA. Comparative molecular modeling of *Anopheles gambiae* CYP6Z1, a mosquito P450 capable of metabolizing DDT. *Proc Natl Acad Sci* 2008;**105**:8855−60.
74. Mitchell SN, Stevenson BJ, Müller P, Wilding CS, Egyir-Yawson A, Field SG, et al. Identification and validation of a gene causing cross-resistance between insecticide classes in *Anopheles gambiae* from Ghana. *Proc Natl Acad Sci USA* 2012;**109**:6147−52.
75. Stevenson BJ, Bibby J, Pignatelli PM, Muangnoicharoen S, O'Neill PM, Liand L-Y, et al. Cytochrome P450 6M2 from the malaria vector *Anopheles gambiae* metabolizes pyrethroids: sequential metabolism of deltamethrin revealed. *Insect Biochem Mol Biol* 2011; **41**:492−502.
76. Riveron JM, Irving H, Ndula M, Barnes KG, Ibrahim SS, Paine MJI, et al. Directionally selected cytochrome P450 alleles are driving the spread of pyrethroid resistance in the major malaria vector *Anopheles funestus*. *Proc Natl Acad Sci* 2013;**110**:252−7.
77. Chandor-Proust A, Bibby J, Régent-Kloeckner M, Roux J, Guittard-Crilat E, Poupardin R, et al. The central role of mosquito cytochrome P450 CYP6Zs in insecticide detoxification revealed by functional expression and structural modelling. *Biochem J* 2013;**455**:75−85.
78. Ibrahim SS, Riveron JM, Stott R, Irving H, Wondji CS. The cytochrome P450 CYP6P4 is responsible for the high pyrethroid resistance in knockdown resistance-free *Anopheles arabiensis*. *Insect Biochem Mol Biol* 2016;**68**:23−32.
79. Stevenson BJ, Pignatelli P, Nikou D, Paine MJI. Pinpointing P450s associated with pyrethroid metabolism in the dengue vector, *Aedes aegypti*: developing new tools to combat insecticide resistance. *PLoS Negl Trop Dis* 2012;**6**:e1595.
80. Riaz MA, Chandor-Proust A, Dauphin-Villemant C, Poupardin R, Jones CM, Strode C, et al. Molecular mechanisms associated with increased tolerance to the neonicotinoid insecticide imidacloprid in the dengue vector *Aedes aegypti*. *Aquat Toxicol* 2013;**126**: 326−37.
81. Kasai S, Komagata O, Itokawa K, Shono T, Ng LC, Kobayashi M, et al. Mechanisms of pyrethroid resistance in the dengue mosquito vector, *Aedes aegypti*: target site insensitivity, penetration, and metabolism. *PLoS Negl Trop Dis* 2014;**8**:e2948.
82. Hardstone MC, Scott JG. A review of the interactions between multiple insecticide resistance loci. *Pestic Biochem Physiol* 2010;**97**:123−8.
83. Edi CV, Djogbénou L, Jenkins AM, Regna K, Muskavitch MAT, Poupardin R, et al. CYP6 P450 enzymes and ACE-1 duplication produce extreme and multiple insecticide resistance in the malaria mosquito *Anopheles gambiae*. *PLoS Genet* 2014;**10**:e1004236.
84. Pasteur N, Georghiou GP, Iseki A. Variation in organophosphate resistance and esterase activity in *Culex quinquefasciatus* say from California. *Génét Sél Evol* 1984; **16**:271−84.
85. Mouchès C, Magnin M, Bergé JB, de Silvestri M, Beyssat V, Pasteur N, et al. Overproduction of detoxifying esterases in organophosphate-resistant Culex mosquitoes and their presence in other insects. *Proc Natl Acad Sci USA* 1987;**84**:2113−6.
86. Georghiou GP, Pasteur N. Electrophoretic esterase patterns in insecticide-resistant and susceptible mosquitoes. *J Econ Entomol* 1978;**71**:201−5.
87. Mouchès C, Pasteur N, Bergé JB, Hyrien O, Raymond M, Robert de Saint Vincent B, et al. Amplification of an esterase gene is responsible for insecticide resistance in a California Culex mosquito. *Science* 1986;**233**:778−80.

Downloaded From Vet-eBooks.com

88. Vaughan A, Hawkes NJ, Hemingway J. Co-amplification explains linkage disequilibrium of two mosquito esterase genes in insecticide-resistant. *Culex Quinquefasciatus Biochem J* 1997;**325**:359−65.
89. Raymond M, Chevillon C, Guillemaud T, Lenormand T, Pasteur NA. An overview of the evolution of overproduced esterases in the mosquito *Culex pipiens*. *Phil Trans R Soc Lond B* 1998;**353**:1707−11.
90. Guillemaud T, Raymond M, Tsagkarakou A, Bernard C, Rochard P, Pasteur N. Quantitative variation and selection of esterase gene amplification in *Culex pipiens*. *Heredity* 1999;**83**:87−99.
91. Weill M, Berticat C, Raymond M, Chevillon C. Quantitative polymerase chain reaction to estimate the number of amplified esterase genes in insecticide-resistant mosquitoes. *Anal Biochem* 2000;**285**:267−70.
92. Poupardin R, Srisukontarat W, Yunta C, Ranson H. Identification of carboxylesterase genes implicated in temephos resistance in the dengue vector *Aedes aegypti*. *PLoS Negl Trop Dis* 2014;**8**:e2743;
92a. Pocquet N, Darriet F, Zumbo B, Milesi P, Thiria J, Bernard V, et al. Insecticide resistance in disease vectors from Mayotte: an opportunity for integrated vector management. *Parasit Vectors* 2014;**7**:299. http://dx.doi.org/10.1186/1756-3305-7-299.
93. Grigoraki L, Lagnel J, Kioulos I, Kampouraki A, Morou E, Labbé P, et al. Transcriptome profiling and genetic study reveal amplified carboxylesterase genes implicated in temephos resistance, in the Asian tiger mosquito *Aedes albopictus*. *PLoS Negl Trop Dis* 2015;**9**: e0003771.
94. Somwang P, Yanola J, Suwan W, Walton C. Enzymes-based resistant mechanism in pyrethroid resistant and susceptible *Aedes aegypti* strains from northern Thailand. *Parasitol Res* 2011;**109**:531−7.
95. Fournier D, Bride JM, Mouchès C, Raymond M, Magnin M, Bergé JB, et al. Biochemical characterization of the esterases A1 and B1 associated with organophosphate resistance in the *Culex pipiens* complex. *Pestic Biochem Physiol* 1987; **27**:211−7.
96. Cuany A, Handani J, Bergé JB, Fournier D, Raymond M, Georghiou GP, et al. Action of esterase B1 on chlorpyrifos in organophosphate-resistant Culex mosquitos. *Pestic Biochem Physiol* 1993;**45**:1−6.
97. Feyereisen R. Molecular biology of insecticide resistance. *Toxicol Lett* 1995;**82**:83−90.
98. Karunaratne SHPP, Hemingway J, Jayawardena KGI, Dassanayaka V, Vaughan A. Kinetic and molecular differences in the amplified and non-amplified esterases from insecticide-resistant and susceptible *Culex quinquefasciatus* mosquitoes. *Biochem Biophys* 1995;**270**:31124−8.
99. Claudianos C, Russell RJ, Oakeshott JG. The same amino acid substitution in orthologous esterases confers organophosphate resistance on the house fly and a blowfly. *Insect Biochem Mol Biol* 1999;**29**:675−86.
100. Pasteur N, Sinègre G, Gabinaud A. Est-2 and Est-3 polymorphism in *Culex pipiens* L. from southern France in relation to organophosphate resistance. *Biochem Genet* 1981;**19**: 499−508.
101. Raymond M, Berticat C, Weill M, Pasteur N, Chevillon C. Insecticide resistance in the mosquito *Culex pipiens*: what have we learned about adaptation? *Genetica* 2001;**112-113**: 1−10.
102. Labbé P, Sidos N, Raymond M, Lenormand T. Resistance gene replacement in the mosquito *Culex pipiens*: fitness estimation from long term cline series. *Genetics* 2009;**182**: 303−12.

Downloaded From Vet-eBooks.com

103. Chevillon C, Bourguet D, Rousset F, Pasteur N, Raymond M. Pleiotropy of adaptive changes in populations: comparisons among insecticide resistance genes in *Culex pipiens*. *Genet Res* 1997;**70**:195−204.
104. Lenormand T, Bourguet D, Guillemaud T, Raymond M. Tracking the evolution of insecticide resistance in the mosquito *Culex pipiens*. *Nature* 1999;**400**:861−4.
105. Berticat C, Boquien G, Raymond M, Chevillon C. Insecticide resistance genes induce a mating competition cost in *Culex pipiens* mosquitoes. *Genet Res Camb* 2002;**79**:41−7.
106. Berticat C, Duron O, Heyse D, Raymond M. Insecticide resistance genes confer a predation cost on mosquitoes, *Culex pipiens*. *Genet Res* 2004;**83**:189−96.
107. Bourguet D, Guillemaud T, Chevillon C, Raymond M. Fitness costs of insecticide resistance in natural breeding sites of the mosquito *Culex pipiens*. *Evolution* 2004;**58**: 128−35.
108. Duron O, Labbé P, Berticat C, Rousset F, Guillot S, Raymond M, et al. High *Wolbachia* density correlates with cost of infection for insecticide resistant *Culex pipiens* mosquitoes. *Evolution* 2006;**60**:303−14.
109. Ffrench-Constant RH, Anthony N, Aronstein K, Rocheleau T, Stilwell G, Richard H. Cyclodiene insecticide resistance: from molecular to population genetics. *Annu Rev Entomol* 2000;**45**:449−66.
110. McKenzie JA. *Ecological and evolutionary aspects of insecticide resistance*. Austin, Texas, USA: Academic Press; 1996.
111. Tantely ML, Tortosa P, Alout H, Berticat C, Berthomieu A, Rutee A, et al. Insecticide resistance in *Culex pipiens quinquefasciatus* and *Aedes albopictus* mosquitoes from La Reunion Island. *Insect Biochem Mol Biol* 2010;**40**:317−24.
112. Pocquet N, Milesi P, Makoundou P, Unal S, Zumbo B, Atyame C, et al. Multiple insecticide resistances in the disease vector Culex p. quinquefasciatus from Western Indian Ocean. *PLoS One* 2013;**8**:e77855.
113. Anthony N, Unruh T, Ganser D, Ffrench-Constant RH. Duplication of the Rdl GABA receptor subunit gene in an insecticide-resistant aphid, *Myzus persicae*. *Mol Gen Genet* 1998;**260**:165−75.
114. Remnant EJ, Good RT, Schmidt JM, Lumb C, Robin C, Daborn PJ, et al. Gene duplication in the major insecticide target site, *Rdl*, in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 2013;**110**:14705−10.
115. Lund AE. Pyrethroid modification of sodium channel: current concepts. *Pestic Biochem Physiol* 1984;**22**:161−8.
116. Vais H, Williamson MS, Devonshire AL, Usherwood PNR. The molecular interactions of pyrethroid insecticides with insect and mammalian sodium channels. *Pest Manag Sci* 2001;**57**:877−88.
117. Soderlund DM, Knipple DC. The molecular biology of knockdown resistance to pyrethroid insecticides. *Insect Biochem Mol Biol* 2003;**33**:563−77.
118. Dong K, Du Y, Rinkevich F, Nomura Y, Xu P, Wang L, et al. Molecular biology of insect sodium channels and pyrethroid resistance. *Insect Biochem Mol Biol* 2014;**50**: 1−17.
119. Chandre F, Darriet F, Duchon S, Finot L, Richet IP, Manguin S, et al. Modifications of pyrethroid effects associated with kdr mutation in *Anopheles gambiae*. *Med Vet Entomol* 2000;**14**:81−8.
120. Williamson MS, Martinez-Torres D, Hick CA, Devonshire AL. Identification of mutations in the housefly para-type sodium channel gene associated with knockdown resistance (kdr) to pyrethroid insecticides. *Mol Gen Genet* 1996;**252**:51−60.

Downloaded From Vet-eBooks.com

121. Martinez-Torres D, Chevillon C, Bergé JB, Pauron D, Brun-Barale A, Bergé JB, et al. Voltage-dependent Na^+ channels in pyrethroid-resistant *Culex pipiens* L. mosquitoes. *Pestic Sci* 1999;**55**:1012−20.
122. Etang J, Vicente JL, Nwane P, Chouaibou M, Morlais I, Do Rosario VE, et al. Polymorphism of intron-1 in the voltage-gated sodium channel gene of *Anopheles gambiae s.s.* populations from Cameroon with emphasis on insecticide knockdown resistance mutations. *Mol Ecol* 2009;**18**:3076−86.
123. Chandre F, Darriet F, Manguin S, Brengues C, Carnevale P, Guillet P. Pyrethroid cross resistance spectrum among populations of *Anopheles gambiae s.s.* from Cote d'ivoire. *J Am Mosq Control Assoc* 1999;**15**:53−9.
124. Ranson H, Jensen B, Vulule JM, Wang X, Hemingway J, Collins FH. Identification of a point mutation in the voltage-gated sodium channel gene of Kenyan *Anopheles gambiae* associated with resistance to DDT and pyrethroids. *Insect Mol Biol* 2000; **9**:491−7.
125. Brengues C, Hawkes NJ, Chandre F, Mccarroll L, Duchon S, Guillet P, et al. Pyrethroid and DDT cross-resistance in *Aedes aegypti* is correlated with novel mutations in the voltage-gated sodium channel gene. *Med Vet Entomol* 2003;**17**:87−94.
126. Du Y, Nomura Y, Satar G, Hu Z, Nauen R, Yang S, et al. Molecular evidence for dual pyrethroid-receptor sites on a mosquito sodium channel. *Proc Natl Acad Sci USA* 2013; **110**:11785−90.
127. Brooke BD. kdr: can a single mutation produce an entire insecticide resistance phenotype? *Trans R Soc Trop Med Hyg* 2008;**102**:524−5.
128. Xu Q, Wang H, Zhang L, Liu N. Kdr allelic variation in pyrethroid resistant mosquitoes, *Culex quinquefasciatus* (S.). *Biochem Biophys Res Commun* 2006;**345**:774−80.
129. Awolola TS, Oduola AO, Oyewole IO, Obansa JB, Amajoh CN, Koekemoer LL, et al. Dynamics of knockdown pyrethroid insecticide resistance alleles in a field population of *Anopheles gambiae s.s.* in southwestern Nigeria. *J Vector Borne Dis* 2007;**44**:181−8.
130. Reimer L, Fondjo E, Patchoké S, Diallo B, Lee Y, Ng A, et al. Relationship between kdr mutation and resistance to pyrethroid and DDT insecticides in natural populations of *Anopheles gambiae*. *J Med Entomol* 2008;**45**:260−6.
131. Dabiré RK, Diabaté A, Namontougou M, Toé KH, Ouari A, Kengne P, et al. Distribution of pyrethroid and DDT resistance and the L1014F kdr mutation in *Anopheles gambiae s.l.* from Burkina Faso (West Africa). *Trans R Soc Trop Med Hyg* 2009;**103**:1113−20.
132. Ramphul U, Boase T, Bass C, Okedi LM, Donnelly MJ, Muller P. Insecticide resistance and its association with target-site mutations in natural populations of *Anopheles gambiae* from eastern Uganda. *Trans R Soc Trop Med Hyg* 2009;**103**:1121−6.
133. Stump AD, Atieli FK, Vulule JM, Besansky NJ. Dynamics of the pyrethroid knockdown resistance allele in Western Kenyan populations of *Anopheles gambiae* in response to insecticide-treated bed net trials. *Am J Trop Med Hyg* 2004;**70**:591−6.
134. Lynd A, Weetman D, Barbosa S, Egyir Yawson A, Mitchell SN, Pinto J, et al. Field, genetic and modelling approaches show strong positive selection acting upon an insecticide resistance mutation in *Anopheles gambiae s.s. Mol Biol Evol* 2010;**27**:1117−25.
135. Norris LC, Main BJ, Lee Y, Collier TC, Fofana A, Cornel AJ, et al. Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. *Proc Natl Acad Sci* 2015;**112**:201418892.
136. Weill M, Chandre F, Brengues C, Manguin S, Akogbéto MC, Pasteur N, et al. The kdr mutation occurs in the Mopti form of *Anopheles gambiae s.s.* through introgression. *Insect Mol Biol* 2000;**9**:451−5.

Downloaded From Vet-eBooks.com

137. Pinto J, Lynd A, Vicente JL, Santolamazza F, Randle NP, Gentile G, et al. Multiple origins of knockdown resistance mutations in the Afrotropical mosquito vector *Anopheles gambiae*. *PLoS One* 2007;**2**:e1243.
138. N'Guessan R, Corbel V, Akogbéto MC, Rowland MW. Reduced efficacy of insecticide-treated nets and indoor residual spraying for malaria control in pyrethroid resistance area, Benin. *Emerg Infect Dis* 2007;**13**:199−206.
139. Casimiro SLR, Hemingway J, Sharp BL, Coleman M. Monitoring the operational impact of insecticide usage for malaria control on *Anopheles funestus* from Mozambique. *Malar J* 2007;**6**:142.
140. Enayati AA, Hemingway J. Pyrethroid insecticide resistance and treated bednets efficacy in malaria control. *Pestic Biochem Physiol* 2006;**84**:116−26.
141. Weill M, Fort P, Berthomieu A, Dubois MP, Pasteur N, Raymond M. A novel acetylcholinesterase gene in mosquitoes codes for the insecticide target and is non-homologous to the *ace* gene in *Drosophila*. *Proc Biol Sci* 2002;**269**:2007−16.
142. Huchard E, Martinez M, Alout H, Douzery EJP, Lutfalla G, Berthomieu A, et al. Acetylcholinesterase genes within the Diptera: takeover and loss in true flies. *Proc Biol Sci* 2006;**273**:2595−604.
143. Fournier D, Mutéro A. Modification of acetylcholinesterase as a mechanism of resistance to insecticides. *Comp Biochem Physiol* 1994;**108C**:19−31.
144. Weill M, Lutfalla G, Mogensen K, Chandre F, Berthomieu A, Berticat C, et al. Insecticide resistance in mosquito vectors. *Nature* 2003;**423**:423−6.
145. Labbé P, Berthomieu A, Berticat C, Alout H, Raymond M, Lenormand T, et al. Independent duplications of the acetylcholinesterase gene conferring insecticide resistance in the mosquito *Culex pipiens*. *Mol Biol Evol* 2007;**24**:1056−67.
146. Nabeshima T, Mori A, Kozaki T, Iwata Y, Hidoh O, Harada S, et al. An amino acid substitution attributable to insecticide-insensitivity of acetylcholinesterase in a Japanese encephalitis vector mosquito, *Culex tritaeniorhynchus*. *Biochem Biophys Res Commun* 2004;**313**:794−801.
147. Alout H, Berthomieu A, Berticat C. Different amino-acid substitutions confer insecticide resistance through acetylcholinesterase 1 insensitivity in *Culex vishnui* and *Culex tritaeniorhynchus* (Diptera: Culicidae) mosquitoes from China. *J Med Entomol* 2007;**44**:463−9.
148. Alout H, Berthomieu A, Hadjivassilis A, Weill M. A new amino-acid substitution in acetylcholinesterase 1 confers insecticide resistance to *Culex pipiens* mosquitoes from Cyprus. *Insect Biochem Mol Biol* 2007;**37**:41−7.
149. Alout H, Labbé P, Berthomieu A, Pasteur N, Weill M. Multiple duplications of the rare ace-1 mutation F290V in *Culex pipiens* natural populations. *Insect Biochem Mol Biol* 2009;**39**:884−91.
150. Alout H, Djogbénou L, Berticat C, Chandre F, Weill M. Comparison of *Anopheles gambiae* and *Culex pipiens* acetycholinesterase 1 biochemical properties. *Comp Biochem Physiol B Biochem Mol Biol* 2008;**150**:271−7.
151. Alout H, Labbé P, Berthomieu A, Makoundou P, Fort P, Pasteur N, et al. High chlorpyrifos resistance in *Culex pipiens* mosquitoes: strong synergy between resistance genes. *Heredity* 2015;**116**:224−31.
152. Assogba BS, Djogbénou LS, Saizonou J, Milesi P, Djossou L, Djegbe I, et al. Phenotypic effects of concomitant insensitive acetylcholinesterase ($ace\text{-}1^R$) and knockdown resistance (kdr^R) in *Anopheles gambiae*: a hindrance for insecticide resistance management for malaria vector control. *Parasit Vectors* 2014;**7**:548.

Downloaded From Vet-eBooks.com

153. Raymond M, Fournier D, Bride JM, Cuany A, Bergé JB, Magnin M, et al. Identification of resistance mechanisms in *Culex pipiens* (Diptera: Culicidae) from southern France: insensitive acetylcholinesterase and detoxifying oxidases. *J Econ Entomol* 1986;**79**: 1452−8.
154. Bourguet D, Roig A, Toutant JP, Arpagaus M. Analysis of molecular forms and pharmacological properties of acetylcholinesterase in several mosquito species. *Neurochem Int* 1997;**31**:65−72.
155. Lenormand T, Guillemaud T, Bourguet D, Raymond M. Appearance and sweep of a gene duplication: adaptive response and potential for new functions in the mosquito *Culex pipiens*. *Evolution* 1998;**52**:1705−12.
156. Labbé P, Milesi P, Yébakima A, Pasteur N, Weill M, Lenormand T. Gene-dosage effects on fitness in recent adaptive duplications: *ace-1* in the mosquito *Culex pipiens*. *Evolution* 2014;**68**:2092−101.
157. Labbé P, Berticat C, Berthomieu A, Unal S, Bernard C, Weill M, et al. Forty years of erratic insecticide resistance evolution in the mosquito *Culex pipiens*. *PLoS Genet* 2007;**3**: e205.
158. Djogbénou L, Chandre F, Berthomieu A, Dabiré RK, Koffi A, Alout H, et al. Evidence of introgression of the *ace-1R* mutation and of the *ace-1* duplication in west African *Anopheles gambiae s.s. PLoS One* 2008;**3**:e2172. 1−7.
159. Dabiré KR, Diabaté A, Namontougou M, Djogbenou L, Kengne P, Simard F, et al. Distribution of insensitive acetylcholinesterase (*ace-1*^{R}) in *Anopheles gambiae s.l.* populations from Burkina Faso (West Africa). *Trop Med Int Heal* 2009;**14**:396−403.
160. Assogba BS, Djogbénou LS, Milesi P, Berthomieu A, Perez J, Ayala D, et al. An *ace-1* gene duplication resorbs the fitness cost associated with resistance in *Anopheles gambiae*, the main malaria mosquito. *Sci Rep* 2015;**5**:14529.
161. Djogbénou L, Labbé P, Chandre F, Pasteur N, Weill M. *Ace-1* duplication in *Anopheles gambiae*: a challenge for malaria control. *Malar J* 2009;**8**:70.
162. Weetman D, Mitchell SN, Wilding CS, Birks DP, Yawson AE, Essandoh J, et al. Contemporary evolution of resistance at the major insecticide target site gene Ace-1 by mutation and copy number variation in the malaria mosquito *Anopheles gambiae*. *Mol Ecol* 2015;**24**:265−72.
163. Djogbénou LS, Assogba B, Essandoh J, Constant EAV, Makoutodé M, Akogbéto M, et al. Estimation of allele-specific *Ace-1* duplication in insecticide-resistant Anopheles mosquitoes from West Africa. *Malar J* 2015;**14**:507.
164. Després L, Stalinski R, Tetreau G, Paris M, Bonin A, Navratil V, et al. Gene expression patterns and sequence polymorphisms associated with mosquito resistance to *Bacillus thuringiensis israelensis* toxins. *BMC Genomics* 2014;**15**:926.
165. Tetreau G, Alessi M, Veyrenc S, David J, Reynaud S. Fate of *Bacillus thuringiensis subsp. israelensis* in the field : evidence for spore recycling and differential persistence of toxins in leaf litters. *Appl Environ Microbiol* 2012;**78**:8362.
166. Nielsen-Leroux C, Pasteur N, Prètre J, Charles J, Sheikh HB, Chevillon C. High resistance to *Bacillus sphaericus* binary toxin in *Culex pipiens* (Diptera: Culicidae): the complex situation of west Mediterranean countries. *J Med Entomol* 2002;**39**:729−35.
167. Chevillon C, Bernard C, Marquine M, Pasteur N. Resistance to *Bacillus sphaericus* in *Culex pipiens* (Diptera: Culicidae): interaction between recessive mutants and evolution in southern France. *J Med Entomol* 2001;**38**:657−64.
168. Nielsen-Leroux C, Charles JF, Georghiou GP. Resistance in a laboratory population of *Culex quinquefasciatus* (Diptera: Culicidae) to *Bacillus sphaericus* binary toxin is due to a

Downloaded From Vet-eBooks.com

change in the receptor on midgut brush-border membranes. *Eur J Biochem* 1995;**228**: 206−10.
169. Darboux I, Pauchet Y, Castella C, Silva-Filha MH, Nielsen-LeRoux C, Charles J-F, et al. Loss of the membrane anchor of the target receptor is a mechanism of bioinsecticide resistance. *Proc Natl Acad Sci USA* 2002;**99**:5830−5.
170. Darboux I, Charles JF, Pauchet Y, Pauron D. Transposon-mediated resistance to *Bacillus sphaericus* in a field-evolved population of *Culex pipiens* (Diptera: Culicidae). *Cell Microbiol* 2007;**9**:2022−9.
171. Saavedra-Rodríguez K, Urdaneta-Marquez L, Rajatileka S, Moulton M, Flores AE, Fernández-Salas I, et al. A mutation in the voltage-gated sodium channel gene associated with pyrethroid resistance in Latin American *Aedes aegypti*. *Insect Mol Biol* 2007;**16**: 785−98.
172. García GP, Flores AE, Fernández-Salas I, Saavedra-Rodríguez K, Reyes-Solis G, Lozano-Fuentes S, et al. Recent rapid rise of a permethrin knock down resistance allele in *Aedes aegypti* in Mexico. *PLoS Negl Trop Dis* 2009;**3**:e531.
173. Kawada H, Higa Y, Nguyen YT, Tran SH, Nguyen HT, Takagi M. Nationwide investigation of the pyrethroid susceptibility of mosquito larvae collected from used tires in Vietnam. *PLoS Negl Trop Dis* 2009;**3**:e391.
174. Marcombe S, Poupardin R, Darriet F, Reynaud S, Bonnet J, Strode C, et al. Exploring the molecular basis of insecticide resistance in the dengue vector *Aedes aegypti*: a case study in Martinique Island (French West Indies). *BMC Genomics* 2009;**10**:494.
175. Brogdon WG, McAllister JC. Insecticide resistance and vector control. *Emerg Infect Dis* 1998;**4**:605−13.
176. Raymond M, Heckel DG, Scott JG. Interactions between pesticide genes. Model and experiment. *Genetics* 1989;**123**:543−51.
177. Bonnet J, Pennetier C, Duchon S, Lapied B, Corbel V. Multi-function oxidases are responsible for the synergistic interactions occurring between repellents and insecticides in mosquitoes. *Parasit Vectors* 2009;**2**:17.
178. Berticat C, Bonnet J, Duchon S, Agnew P, Weill M, Corbel V. Costs and benefits of multiple resistance to insecticides for *Culex quinquefasciatus* mosquitoes. *BMC Evol Biol* 2008;**8**.
179. Vontas JG, McCarroll L, Karunaratne SHPP, Louis C, Hurd H, Hemingway J. Does environmental stress affect insect-vectored parasite transmission? *Physiol Entomol* 2004; **29**:210−3.
180. Rivero A, Vézilier J, Weill MM, Read AF, Gandon S. Insecticide control of vector-borne diseases: when is insecticide resistance a problem? *PLoS Pathog* 2010;**6**:e1001000.
181. Alout H, Ndam NT, Sandeu MM, Djégbe I, Chandre F, Dabiré RK, et al. Insecticide resistance alleles affect vector competence of *Anopheles gambiae s.s.* for *Plasmodium falciparum* field isolates. *PLoS One* 2013;**8**:e63849.
182. Alout H, Djègbè I, Chandre F, Djogbénou LS, Dabiré RK, Corbel V, et al. Insecticide exposure impacts vector-parasite interactions in insecticide-resistant malaria vectors. *Proc R Soc B* 2014;**281**:20140389.
183. Kelly-Hope LA, Ranson H, Hemingway J. Lessons from the past: managing insecticide resistance in malaria control and eradication programmes. *Lancet Infect Dis* 2008;**8**: 387−9.

Downloaded From Vet-eBooks.com

This page intentionally left blank

Downloaded From Vet-eBooks.com

Genetics of Major Insect Vectors

15

P.L. Dorn[1], S. Justi[2], E.S. Krafsur[3], G.C. Lanzaro[4], A.J. Cornel[4,5], Y. Lee[4], C.A. Hill[6]

[1]Loyola University New Orleans, New Orleans, LA, United States; [2]University of Vermont, Burlington, VT, United States; [3]Iowa State University, Ames, IA, United States; [4]University of California at Davis, Davis, CA, United States; [5]Mosquito Control Research Lab, Parlier, CA, United States; [6]Purdue University, West Lafayette, IN, United States

1. Introduction

1.1 Significance and Control of Vector-Borne Disease

Vector-borne diseases are major contributors to the global disease burden. They are responsible for >17% of all infectious disease and 1 million deaths annually[1] (Fig. 15.1). Control of insect vectors is often the best, and sometimes the only, way to protect humans from these destructive diseases. Vector control is a moving target with globalization and demographic changes causing changes in infection patterns (e.g., rapid spread, urbanization, and appearance in nonendemic countries); and the current unprecedented degradation of the global environment is affecting rates and patterns of vector-borne disease in ways that are still largely unknown. It is increasingly apparent that effective vector control requires multidisciplinary, community-based, and environmentally sustainable approaches that are responsive to local conditions, such as the Ecohealth approach, that has been successfully applied to Chagas disease in Central America.[2]

1.2 Contributions of Genetic Studies of Vectors to Understanding Disease Epidemiology and Effective Disease Control Methods

Within this multidisciplinary context, studies of vector genetics have a major role in clarifying vector-borne disease epidemiology and designing successful control methods. Phylogenetic analyses of major species have helped identify new species, subspecies, and cryptic species, which, in conjunction with ecological studies, have implicated epidemiologically important taxa, targets for control. Cytogenetics has revealed the role of the evolution of chromosome structure in insect vector speciation. Population genetic studies have uncovered the complex population structures of insect vector populations and *gene flow* among populations revealing the geographical coverage needed for control and the source of reinfesting insects. Genetic control methods, such as the sterile insect technique,[3] or introduction of refractory traits or transgenic symbionts carrying molecules toxic to pathogens, can add to the arsenal.

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00015-9
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

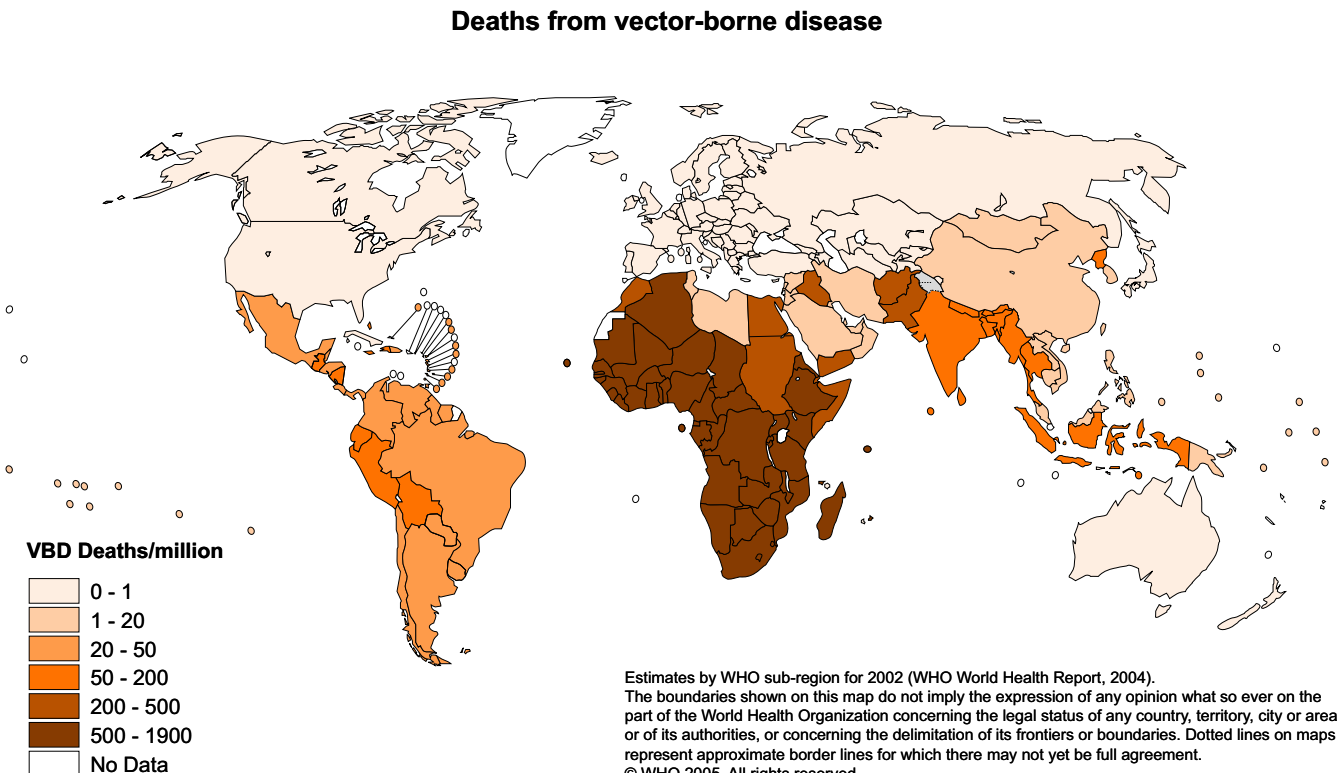


Figure 15.1 Estimates of vector-borne disease deaths per million inhabitants, with permission, copyright WHO. Available from: http://www.who.int/heli/risks/vectors/vector/en/index.html.

Downloaded From Vet-eBooks.com

Molecular genetics and new genome and proteome tools promise advances in understanding the genetic basis of vector capacity including habitat and host preference, innate immunity, drought tolerance, insecticide resistance, among other phenomena, and the development of new attractants and repellents.

2. Genetics of Tsetse Flies and African Trypanosomiasis

2.1 Introduction

Tsetse flies (Diptera: Glossinidae) (pronounced "tsee-tsee," Fig. 15.2) are among the most important insects in sub-Saharan Africa because they are obligate blood feeders and the vectors of African trypanosomiasis caused by hemoflagellate trypanosomes that kill humans and domestic mammals. More than 70 million people are at risk for human African trypanosomiasis (HAT) in 36 countries.[4] Nagana, animal trypanosomiasis (AAT), was estimated to cost African agriculture US $4.5 billion per year[5] via loss of food, dung, and drafting power. Additional reviews include.[6–9]

2.2 The Family Glossinidae

Tsetse flies are assigned to the family Glossinidae with extant tsetse flies classified into a single genus, *Glossina* Wiedemann 1830, containing four subgenera, *Machadomyia*, *Austenina*, *Nemorhina*, and *Glossina* that correspond to the Fusca (forest), Palpalis (river and lake), and Morsitans (savanna) species groups, respectively.[9] Subgenus *Machadomyia* consists only of two *G. austeni* subspecies. Thirty-four taxa have been described, consisting of 23 species and 7 species complexes of 17 named subspecies that differ slightly morphologically, if at all, and are mostly *allopatric*. There is an



Figure 15.2 Resting tsetse.
Photo courtesy of the DFID Animal Health Program.

Downloaded From Vet-eBooks.com

extinct sister group to the Glossinidae known from the Florissant shale of Colorado, and similar tsetse-like fossils were uncovered in Oligocene strata in Germany indicating formerly a much wider geographical distribution.

Three species complexes are geographically widespread and of much medical and economic importance (Figs. 15.3−15.5). The most thoroughly examined is *Glossina morsitans* s.l. and its close relative, *Glossina swynnertoni*. *G. morsitans* s.l. comprises *G. morsitans morsitans, G. morsitans centralis*, and *G. submorsitans*. Genetic data suggest longstanding reproductive isolation. *Glossina palpalis* s.l. comprises *G. palpalis palpalis* and *G. palpalis gambiense*; studies suggest *incipient speciation* in *G. p. palpalis*[10,11] and *G. p. gambiense*.[12] The foregoing taxa are *allopatric* and hybrid males are sterile, the females typically sterile or semisterile.[13] Based on morphological criteria, *Glossina fuscipes* s.l. consists of allopatric *G. fuscipes fuscipes, G. fuscipes martini*, and *G. fuscipes quanzensis*.[14] Dyer et al.[15] however, found insufficient genetic evidence of cryptic speciation among the *fuscipes* subspecies. Further work is necessary to sort out the taxonomic status of Morsitans and Palpalis group taxa.

2.3 Genetics and Population Genetics of Tsetse Flies

2.3.1 Cytogenetics

All tsetse flies examined cytologically have two pairs of metacentric autosomes and a sex bivalent: $2N = 4 + XY$. Many also have heterochromatic supernumerary chromosomes, and sex chromosome polymorphisms have been recorded in wild *G. p. palpalis*. Taxa within *Morsitans* and *Palpalis* flies can be separated by *pericentric* and *paracentric chromosome inversions*.

2.3.2 Genetic Variability Based on Microsatellite Loci and mtDNA

Microsatellite diversities, averaged over loci, varied from a low 0.43 in *G. f. fuscipes* to 0.81 in *G. m. submorsitans* (Table 15.1), with a lower mean than that of the housefly, *Musca domestica*. Again, lower genetic variation is likely due to a smaller *effective population size* in the tsetse flies. Cytochrome oxidase subunit I and ribosomal 16S reveal many sequence variants in large samples of Morsitans and Palpalis group flies (Table 15.2). Mitochondrial diversity was least in *G. swynnertoni* (found in a small region of northcentral Tanzania).

Low diversities in *G. m. centralis* and southern African *Glossina pallidipes* reflect earlier demographic events including the 19th century rinderpest epizootic that virtually eliminated the Morsitans group flies.[16] In *G. pallidipes*, microsatellite and mitochondrial diversities were less in southern Africa than in East Africa and both were strongly correlated with each other; this variation was consistent with a severe and prolonged reduction in population sizes in southern Africa.

Downloaded From Vet-eBooks.com

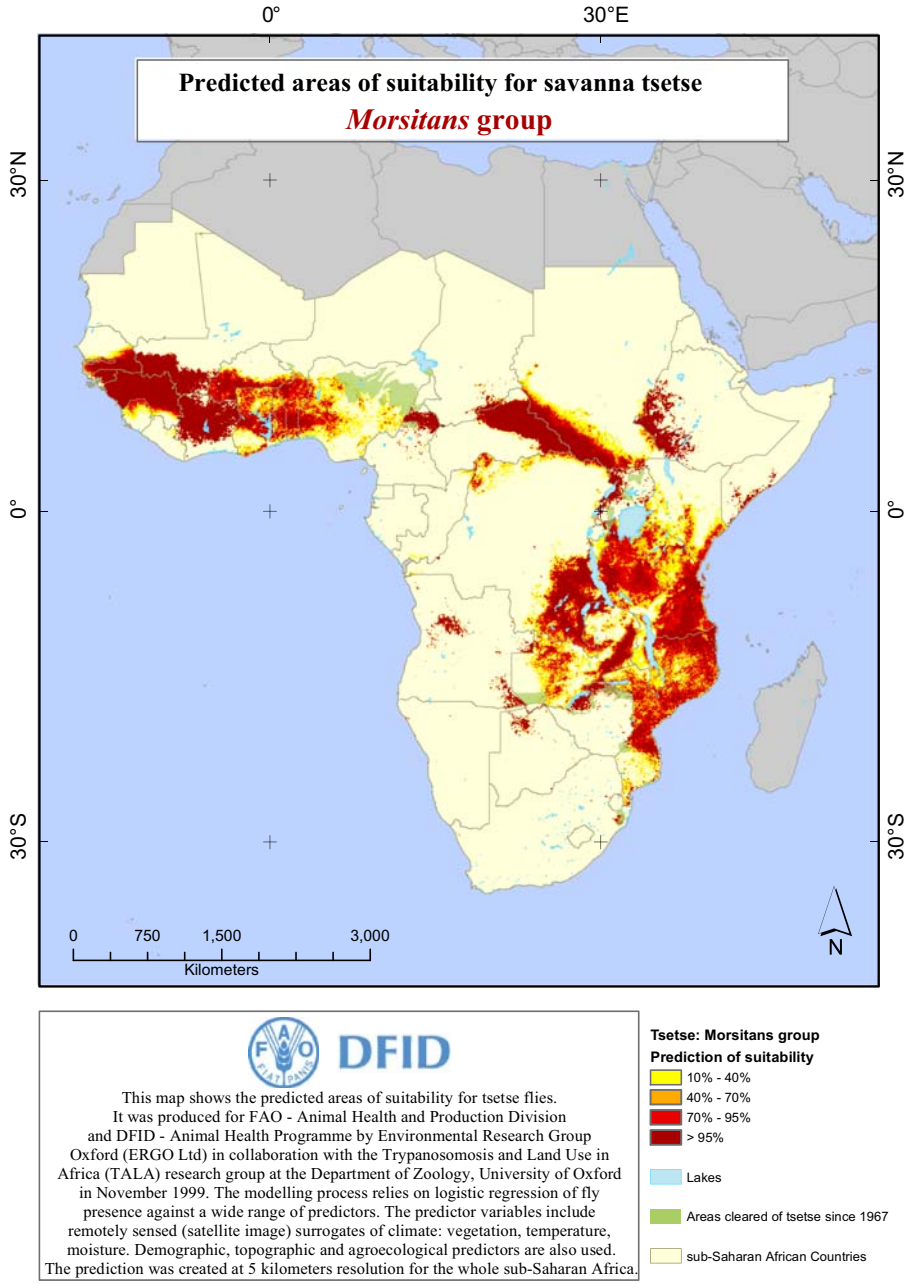


Figure 15.3 Predicted distribution of Morsitans group tsetse flies in Africa. Reproduced with permission from Wint W, Rogers D. Food and Agriculture Organization of the United Nations; 2000. http://www.fao.org/ag/againfo/programmes/en/paat/maps.html.

Downloaded From Vet-eBooks.com

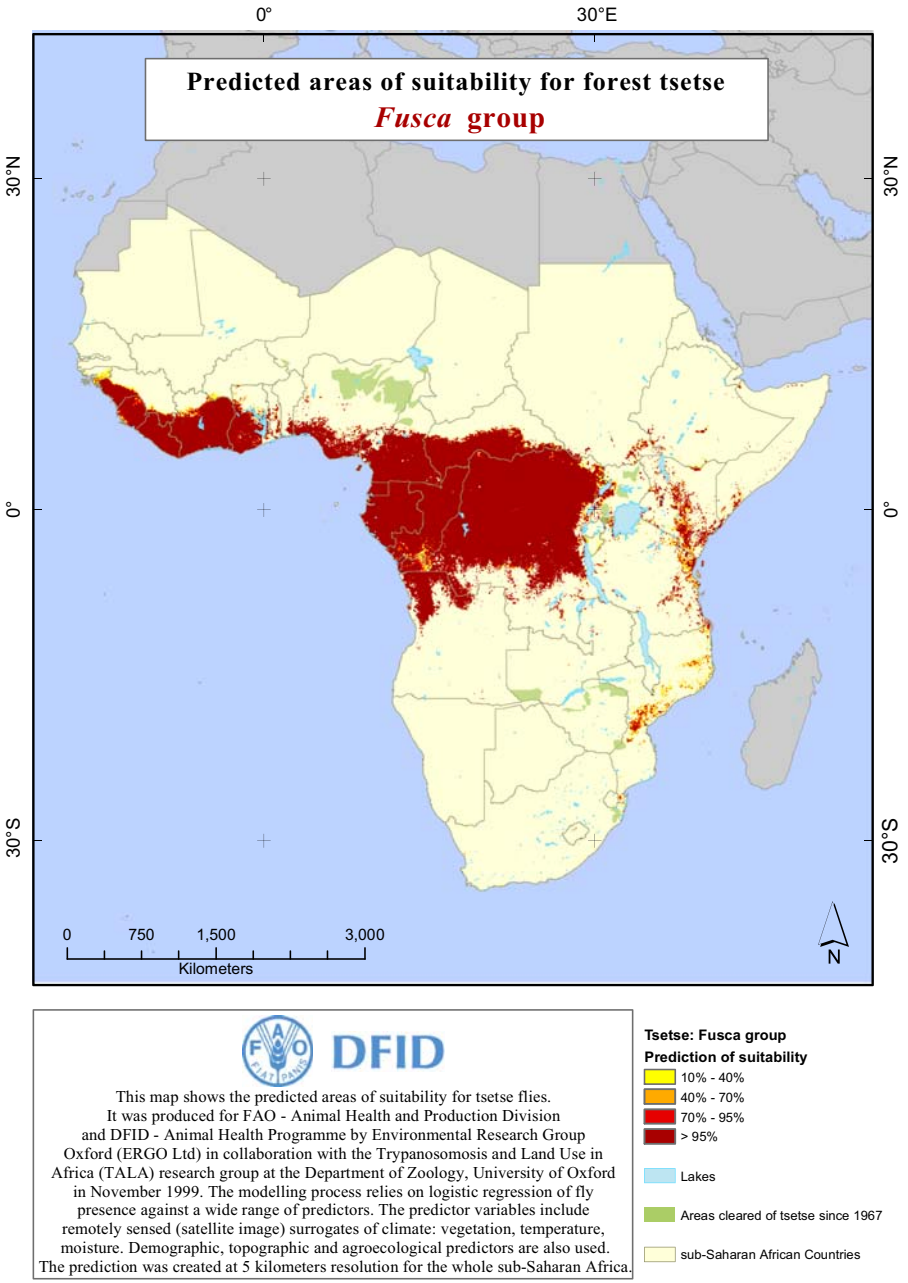


Figure 15.4 Predicted distribution of Fusca group tsetse flies in Africa. Reproduced with permission from Wint W, Rogers D. Food and Agriculture Organization of the United Nations; 2000. http://www.fao.org/ag/againfo/programmes/en/paat/maps.html.

Downloaded From Vet-eBooks.com

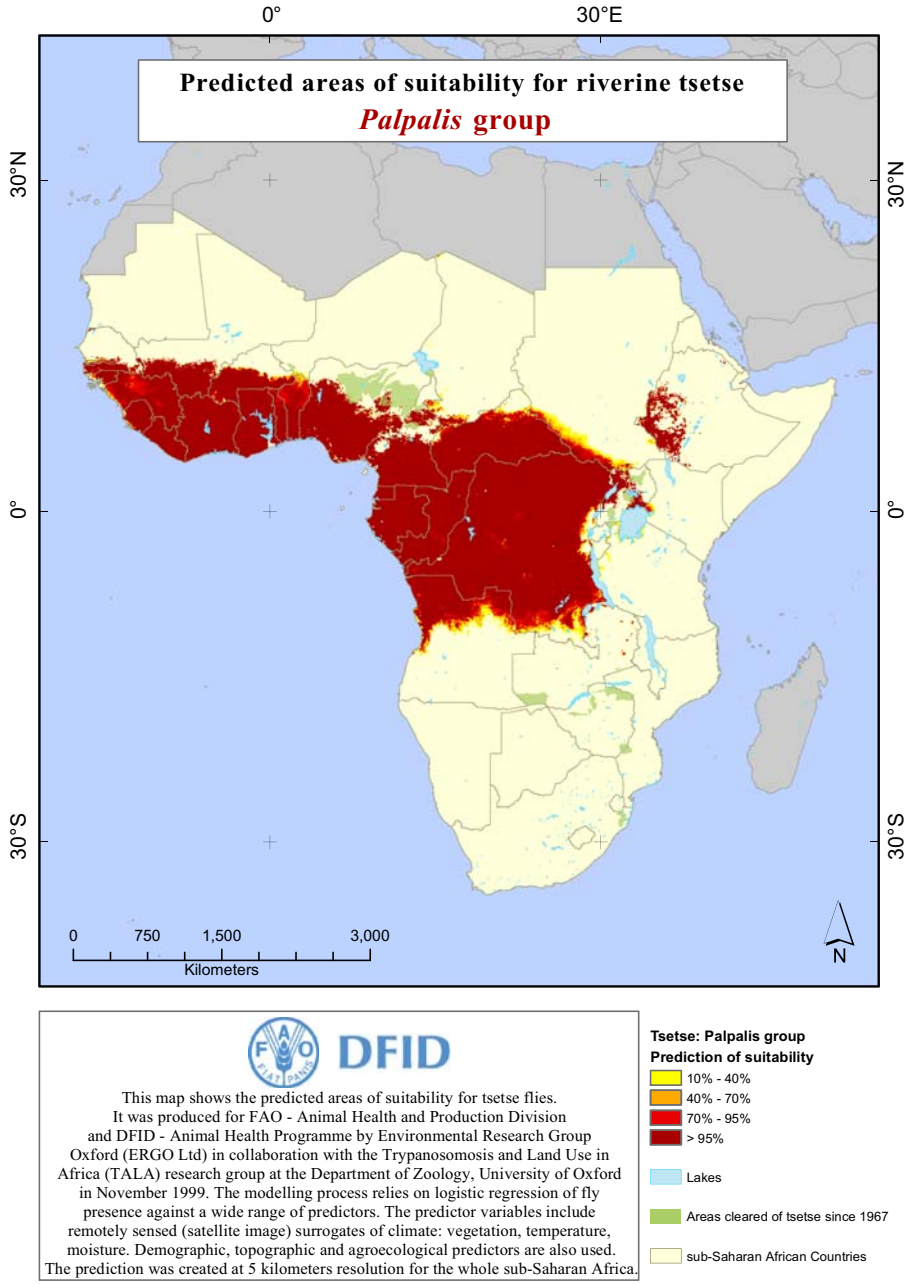


Figure 15.5 Predicted distribution of Palpalis group tsetse flies in Africa. Reproduced with permission from Wint W, Rogers D. Food and Agriculture Organization of the United Nations; 2000. http://www.fao.org/ag/againfo/programmes/en/paat/maps.html.

Downloaded From Vet-eBooks.com

Table 15.1 Microsatellite Diversities and Tests for Random Matings *F* in *Glossina* spp. and the Housefly

	No. Populations	No. Loci	Alleles per Locus	Diversity H_e	Within Demes F_{IS}	Among Demes F_{ST}
G. m. morsitans	6	6	11.0 ± 5.6	0.73 ± 0.06	0.03 ± 0.03	0.19 ± 0.05
G. m. morsitans	9	7	28.6 ± 4.5	0.74 ± 0.05	0.17 ± 0.07	0.13 ± 0.01
G. m. centralis	7	7	8.8 ± 3.7	0.70 ± 0.09	-0.12 ± 0.04	0.19 ± 0.04
G. m. submorsitans	7	7	12.7 ± 6.2	0.81 ± 0.04	0.03 ± 0.03	0.17 ± 0.07
G. pallidipes	21	8	26.8 ± 8.7	0.80 ± 0.03	0.07 ± 0.03	0.18 ± 0.02
G. f. fuscipes[a]	8	5	8.2 ± 3.6	0.43 ± 0.07	0.11 ± 0.05	0.22 ± 0.07
Musca domestica	14	7	7.9 ± 1.1	0.86 ± 0.02	0.08 ± 0.02	0.13 ± 0.02

[a] Based on data from Abila PP, Slotman MA, Parmakelis A, Dion KB, Robinson AS, Muwanika VB, et al. High levels of genetic differentiation between Ugandan *Glossina fuscipes fuscipes* populations separated by Lake Kyoga. *PLoS Neglected Tropical Diseases* 2008;**2**(5):e242.

Downloaded From Vet-eBooks.com

Table 15.2 Mitochondrial Diversities and Genetic Differentiation in Wild *Glossina* Species

	Method	No. Populations	No. Flies	No. Haplotypes	Haplotype Diversity, H_S	F_{ST}
G. m. morsitans	SSCP	5	111	25	0.81 ± 0.04	0.09 ± 0.02
G. m. morsitans	ABI 3730	7	96	33	0.81	0.40 ± 0.08
G. m. centralis	SSCP	6	265	7	0.54 ± 0.16	0.81 ± 0.07
G. m. submorsitans	SSCP	7	282	26	0.51 ± 0.12	0.35
G. pallidipes	SSCP	21	624	39	0.42 ± 0.02	0.52 ± 0.001
G. pallidipes	ABI 3730	23	873	181	0.73 ± 0.09	0.47 ± 0.07
G. p. gambiensis	SSCP	13	372	9	0.18	0.68
G. swynnertoni	ABI 3730	8	149	18	0.59 ± 0.10	0.04 ± 0.003
G. f. fuscipes[a]	ABI 3730	22	284	36	0.91 ± 0.008	0.60 ± 0.07

[a] Based on data from: Beadell JS, Hyseni C, Abila PP, Azabo R, Enyaru JC, Ouma JO, et al. Phylogeography and population structure of *Glossina fuscipes fuscipes* in Uganda: implications for control of tsetse. *PLoS Neglected Tropical Diseases* 2010;**4**(3):e636.

Downloaded From Vet-eBooks.com

2.3.3 Population Structure of Tsetse Flies

Random genetic drift was pronounced in all taxa, *G. morsitans* s.l., *G. pallidipes*, *G. swynnertoni*, *G. f. fuscipes*, *G. p. palpalis*, and *G. p. gambiense*, leading to highly significant levels of genetic differentiation among conspecific populations. Most population samples were differentiated even when within 25−50 km of each other, and genetic diversity in laboratory cultures was only mildly attenuated compared with their field cousins, with the possible exception of mitochondrial diversity *H*s (8 haplotypes, $Hs = 0.36$) in a longstanding *G. austeni* culture. Mean estimates of F_{ST} indicate low levels of gene flow (Tables 15.1 and 15.2) and for most tsetse taxa, the mean numbers of reproductive flies exchanged among populations is generally less than one or two per generation, indicative of strong genetic drift.

Strong genetic drift is surprising in light of high dispersion, as shown by mark-recapture studies.[17−19] Spatial variations in natural selection, such as temperature and moisture conditions, govern the distribution of tsetse flies, and empirical evidence that spatially separated demes have adapted to their different environments may provide an explanation.[20]

2.4 Tsetse Population Management

There are no vaccines for HAT, and pharmaceutical treatment is expensive, dangerous, and unavailable to most people at risk, thus it is best controlled by eliminating its insect vectors. Older methods of tsetse population management have failed due to invasion from nearby, untreated populations. Genetic methods, such as SIT, have been applied experimentally to several tsetse taxa,[9,21] and laboratory experiments and simulations have demonstrated the hypothetical efficacy of using *cytoplasmic incompatibility* conferred by *Wolbachia* together with transgenic gut symbionts as a means of driving trypanosome-refractory phenotypes into natural populations.[22]Application of SIT on the African continent is not recommended because of its very high cost, poor sterile fly competitiveness, and the availability of proven cost-effective methods,[19] which is confirmed by simulation models.[23] Replacement of natural vector populations with conspecific nonvectors may ultimately prove impractical due to financial costs of field application and follow-up.

2.5 Further Work Needed

Barriers to developing further scientific knowledge of tsetse fly biology include a severe paucity of laboratory cultures representative of natural populations. The pronounced genetic variation among natural tsetse populations argues for geographically more extensive sampling across the geographical range to assess genetic variation and reciprocal crossing of different lines to assay fertilities and uncover additional sibling species. Vector−parasite coadaptations have important epidemiological and economic consequences, but how they vary spatially is unknown.

Regarding tsetse fly population management, it is noteworthy that their historical distribution and abundance is unchanged except for their elimination in relatively

Downloaded From Vet-eBooks.com

small areas on the northern and southern margins—southwestern Zambia, northeastern Zimbabwe, northern Nigeria, and the Okavango in Botswana. Proposed genetic approaches to trypanosomiasis control are interesting, and related research is yielding important scientific insights, but laboratory elegance alone would seem unlikely to overcome the dynamic nature of tsetse fly populations in their natural habitats.

The breeding structures of *Glossina brevipalpis*, *G. f. quanzensis*, and *Glossina longipalpis* are unknown and these are the vectors in Mozambique and much of Central Africa. The present view that most, if not all, Morsitans and many Palpalis group populations are local may serve to define areas in which systematic vector management schemes may be applied without massive immigration from untreated, conspecific populations. Effective and affordable genetically based area-wide tsetse fly population management is unlikely to be developed in the foreseeable future while coordinated application of conventional methods can achieve highly effective control of *Glossina* and AAT.[24]

3. Genetics of the Triatominae (Hemiptera, Reduviidae) and Chagas Disease

3.1 Introduction

Chagas disease, a zoonosis caused by the flagellate protozoan *Trypanosoma cruzi* (Kinetoplastea: Trypanosomatida), is among the most serious neglected tropical diseases in Latin America. Although rarely fatal in its early acute stage, in about 30% of those infected it progresses to a debilitating chronic disease that involves severe cardiac and intestinal lesions, usually fatal. There is no vaccine available, and treatment is woefully inadequate.

Chagas disease, also called American trypanosomiasis, ranges from the southern United States to Argentinean Patagonia, and human disease from Mexico to northern Argentina, mostly in poor, rural areas where houses are infested with insect vectors belonging to the subfamily Triatominae (Hemiptera, Reduviidae) (Fig. 15.6). With deforestation and migration, Chagas is increasingly found in urban areas and even in nonendemic countries. Regional Chagas control initiatives have resulted in a dramatic reduction of disease prevalence due to decreased vector transmission and an increase in blood donation screening. Despite these gains, nearly 6 million people remain infected.[25]

3.2 Chagas Disease Vectors

Chagas disease vectors (kissing or conenose bugs) are insects from the subfamily Triatominae and the only proven vectors of American trypanosomiasis. They require a blood meal to molt and lay eggs, and acquire the Chagas parasite when they feed on a *T. cruzi*−infected mammal. Triatomines transmit the parasite to a new host via parasite-contaminated feces deposited on the skin or mucous membranes of the new

Downloaded From Vet-eBooks.com

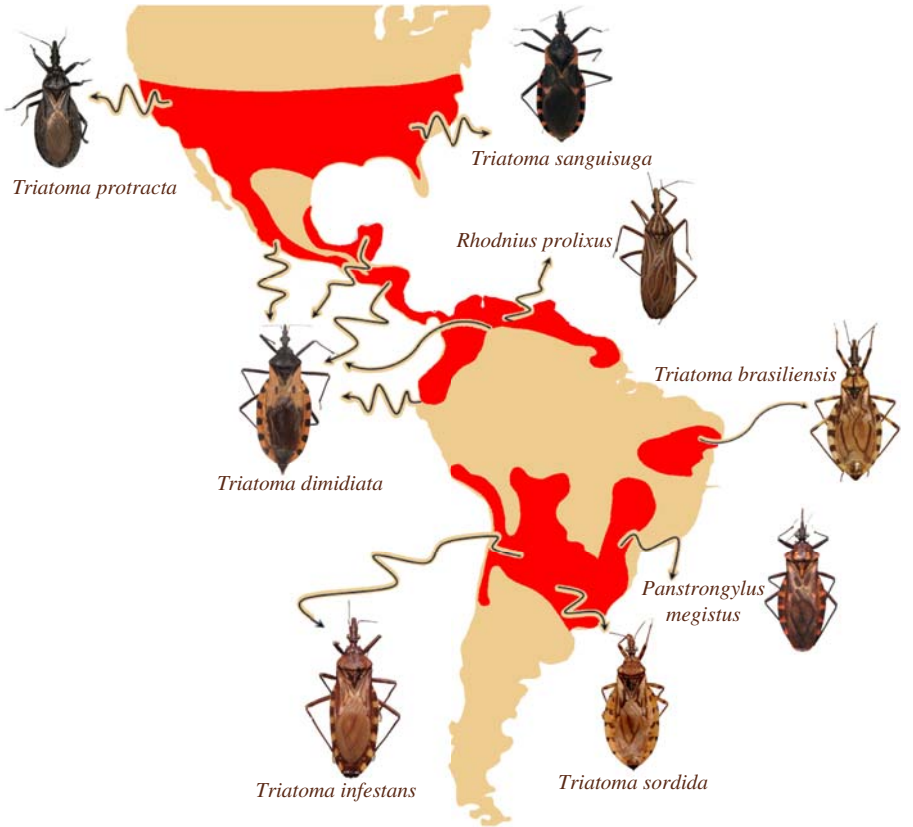


Figure 15.6 Approximate distribution of triatomine species of major epidemiological relevance.
Map adapted by permission from Macmillan Publishers Ltd: Heredity 108:190−202, copyright (2011). Photos: *R. prolixus*, P. *megistus*, *T. brasiliensis*, *T. infestans* and *T. sordida*: Laboratório Nacional e Internacional de Referência em Taxonomia de Triatomíneos (Oswaldo Cruz Institute, FIOCRUZ, Brazil); *T. dimidiata* and *T. sanguisiga*: H. Baquet; *T. protracta*: J. Schmidt.

host during a blood meal. Domesticity is a key determinant of vector capacity, and less than a dozen of the over 140 Triatominae species are known to be well adapted to human dwellings, being the major vectors.

3.3 Evolution of the Triatominae

Triatominae is a subfamily of Reduviidae (assassin bugs), which are mostly predators of other arthropods, while triatomines have evolved hematophagy of nest-dwelling vertebrates. It is unclear whether this hematophagy arose once (*monophyletic* origin)[26–28] (Fig. 15.7), or several times (paraphyletic/*polyphyletic* origin)[29,30] within the subfamily. Studies including species from the Alberprosenini, Bolboderini, and Cavernicolini tribes are needed to fill gaps in the comprehensive picture of the evolution of the Triatominae.

Downloaded From Vet-eBooks.com

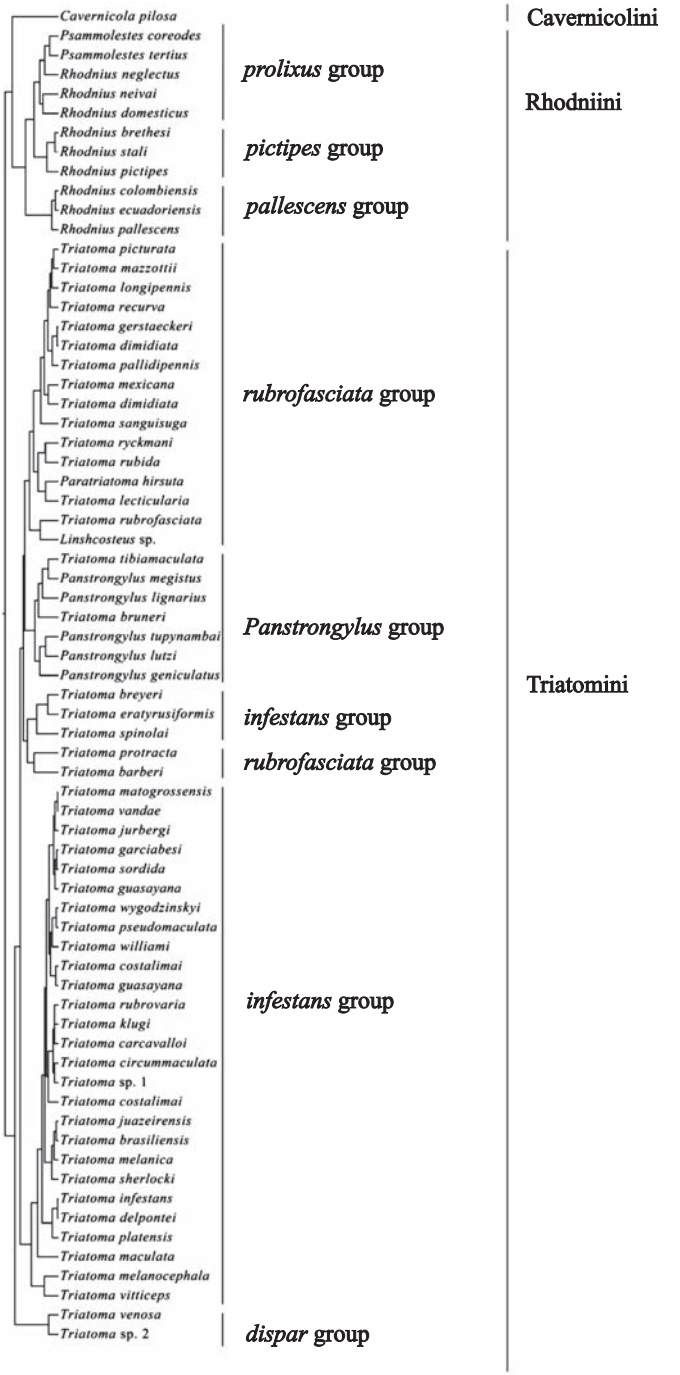


Figure 15.7 Bayesian phylogeny of the Triatominae.[28] *Triatoma* sp. 1 is an undescribed taxon, morphologically similar to *T. gasayana*, with short wings, found by F. Noireau in the Bolivian Chaco. *Triatoma* sp. 2 is an undescribed taxon collected from mammal dwelling in Sucumbios, Ecuador.[30]

Downloaded From Vet-eBooks.com

3.3.1 The Five Triatominae Tribes

Triatominae are classified into 5 tribes and 15 genera and include 147 described species[31] (Table 15.3); most (~125) occur exclusively in the New World. However, members of the *rubrofasciata* complex are worldwide, likely spread on ships, and a few others are found only in Asia. The species most important for human transmission are: *Triatoma brasiliensis* and *Triatoma infestans* in South America, *Triatoma dimidiata* and *Rhodnius prolixus* in Central America and northern South America (Fig. 15.6), which belong to the tribes Triatomini and Rhodniini.

The Triatomini tribe is the most diverse, with over 100 described species including *Triatoma* and *Panstrongylus*, the most diverse and epidemiologically important (Table 15.3). Few morphological differences separate the Triatomini genera[32] and some rank assignments are still unresolved[31,33]; no cladistic analysis is available to date. *T. dimidiata* is the most important vector in southern Mexico, Central America, and a secondary vector in northern South America (Fig. 15.6). *Triatoma infestans* and *T. brasiliensis* are important vectors across South America and in Brazil, respectively (Fig. 15.6).

Two genera comprise the Rhodniini tribe: *Rhodnius* and *Psammolestes* (Table 15.3). Known molecular phylogenies include *Psammolestes* within the *Rhodnius* clade[34] despite its morphological adaptations associated with living in bird nests. *Rhodnius* is morphologically quite different from the other Triatominae, especially in the head morphology; many species are difficult to distinguish.[35] *Rhodnius prolixus* is currently the main Chagas vector in Colombia and Venezuela and appears to have been introduced (and now eliminated from) Central America.[36]

3.4 Cytogenetics of the Triatominae

Most Triatominae species have a diploid nuclear genome of 20 autosomes and XY sex chromosomes.[37] The number of sex chromosomes varies across the species, due to the fragmentation of the X-chromosome in some species into two or three fragments (Fig. 15.8), and the amount and localization of heterochromatin varies considerably among and even within some species. For example, Andean *T. infestans* specimens present 50% more C-banded heterochromatin and 30% more DNA content than non-Andean,[37] which has led to the suggestion that the Bolivian Andean valleys are the origin of *T. infestans*. Chromosomal differences were also important in supporting a *cryptic species* in *T. dimidiata*.[38] The diversity of triatomine chromosome structure makes it a model system for understanding chromosome structure and its evolution, and the role of heterochromatin.

3.5 Genetic Diversity of the Triatominae

In general, *T. infestans* and *R. prolixus* populations show less genetic diversity than *T. dimidiata* and *T. brasiliensis* (Table 15.4). A diminished genetic repertoire in largely domestic populations (*T. infestans* and *R. prolixus*) would be predicted due to founder effects and genetic drift in isolated populations[39] and greater genetic variability in

Downloaded From Vet-eBooks.com

Table 15.3 Updated List of Triatominae Described Species and Group Assignment Based[166]

Tribes	Genera	Group	Complex	Subcomplex	Species	Number of Species
Alberprosenini	*Alberprosenia*				*goyovargasi, malheiroi*	2
Bolboderini	*Belminus*				*corredori, costaricencis, ferroae, herreri, laportei, peruvianus, pittieri, rugulosus*	8
	Bolbodera				*scaborsa*	1
	Microtriatoma				*borbai, trinidadensis*	2
	Parabelminus				*carioca, yurupucu*	2
Cavernicolini	*Cavernicola*				*lenti, pilosa*	2
Rhodniini	*Pasmmolestes*				*arthuri, coreodes, tertius*	3
	Rhodnius	*prolixus*			*barretti, dalessandroi, domesticus, milesi, montenegrensis, nasutus, neglectus, neivai, prolixus, robustus*	10
		pictipes			*amazonicus, brethesi, paraensis, pictipes, stali, zeledoni*	6
		pallescens			*colombiensis, ecuadoriensis, pallescens*	3
Triatomini	*Dipetalogaster*				*maxima*	1
	Eratyrus				*cuspidatus, mucronatus*	2
	Hermanlentia				*matsunoi*	1
	Linshcosteus				*carnifez, chota, confumus, costalis, kali, karupus*	6

Continued

Downloaded From Vet-eBooks.com

Table 15.3 Updated List of Triatominae Described Species and Group Assignment Based[166]—cont'd

Tribes	Genera	Group	Complex	Subcomplex	Species	Number of Species
	Panstrongylus				*chinai, diasi, geniculatus, guentheri, howardi, humeralis, lenti, lignarius, lutzi, megistus, mitarakaensis, rufotuberculatus, sherlocki, tupynambai*	14
	Paratriatoma				*hirsuta*	1
	Triatoma	*rubrofasciata*	*phyllosoma* (*Meccus*)	*dimidiata*	*dimidiata, hegneri, brailovskyi, gomeznunezi*	4
				phyllosoma	*bassolsae, bolivari, longipennis, mazzottii, mexicana, pallidipennis, phyllosoma, picturata, ryckmani,*	9
			flavida (*Nesotriatoma*)		*flavida, bruneri, obscura*	3
			rubrofasciata		*amicitiae, bouvieri, cavernicola, leopoldi, migrans, pugasi, rubrofasciata, sinica*	8
			protracta		*barberi, incrassata, neotomae, nitida, peninsularis, protracta, sinaloensis*	7
			lecticularia		*gerstaeckeri, indictiva, lecticularia, recurva, rubida, sanguisuga*	6

Downloaded From Vet-eBooks.com

		dispar	*dispar*		*bolviana, carrioni, dispar, nigromaculata, venosa*	5
		infestans	*infestans*	*brasiliensis*	*brasiliensis, juazeirensis, melanica, melanocephala, petrochiae, lenti, sherlocki (tibiamaculata?) (vitticeps?)*	9
				infestans	*delpontei, infestans, platensis*	3
				maculata	*arthurneivai, maculata, pseudomaculata, wygodzinskyi*	4
				matogrossensis	*baratai, costalimai, deaneorum, guazu, jatai, jurbergi, matogrossensis, vandae, williami*	9
				rubrovaria	*carcavalloi, circummaculata, klugi, limai, oliveirai, pintodiasi, rubrovaria*	7
				sordida	*garciabesi, guasayana, patagonica, sordida*	4
			spinolai (*Mepraia*)		*breyeri, eratyrusiformis, gajardoi, parapatrica, spinolai*	5
						147

Downloaded From Vet-eBooks.com



Figure 15.8 Current knowledge of Triatominae karyotypes across the diversity mapped onto a Bayesian phylogeny of the Triatominae indicating that multiple X-chromosomes have appeared several times during Triatominae evolution. *Colors* indicate number of chromosomes; *no color*, information not available. The reader is referred to the online version for color.

Downloaded From Vet-eBooks.com

Table 15.4 Mitochondrial and Nuclear Diversities in Triatomine Populations

Region	Species	Marker	No. Populations	n	No. Haplotypes	Haplotype Diversity H_d	References
North and Central America	*T. dimidiata*[a]	*cyt b*	12	24	21	0.960	167
	T. dimidiata	*cyt b*	7	58	15	0.901	50
	T. sanguisuga	*cyt b*	1	54	37	0.978	168
	T. dimidiata[a]	*ITS2*	93	190	39	0.918	169,167
	T. dimidiata	*ITS2*	7	58	1	0	50
South America	*T. dimidiata*	*COI*	12	82	63	0.985	51
	T. dimidiata	*ND4*	22	228	155	0.991	52
	T. infestans	*cyt b*	43	98	11	0.737	170,171
	T. infestans	*cyt b*	20	211	19	0.727	40
	R. prolixus	*cyt b*	34	551	15	0.518	42
	T. brasiliensis	*cyt b*	4	361	29	0.905	172
	T. brasiliensis[b]	*cyt b*	17	136	35	0.920	173
	T. infestans	*ITS2*	31	35	5	0.591	174
	T. infestans	*ITS2*	20	193	3	0−0.476	40

[a]May include a cryptic species.
[b]Includes three proposed species: *T. brasiliensis/macromelasoma*, *T. juazeirensis*, and *T. melanica.*

Downloaded From Vet-eBooks.com

populations found in diverse habitats (*T. dimidiata* and *T. brasiliensis*), due to distinct selective pressures in each habitat. Surprisingly, sylvan *T. infestans*[40,41] and *R. prolixus*[42] show lower, or comparable genetic diversity with that of domestic populations. The strong selective pressure of insecticide application can result in diminished,[43] an increased, or no effect on the genetic variability of a population.[44,45]

3.6 Population Structure of the Triatominae

Active (flying up to 2 km[46]) and passive transport (by human activity and perhaps migratory birds) have been important in spreading and mixing triatomine populations. Overall, studies in *T. dimidiata* and *T. infestans* show that at larger geographical scales (populations >50 km apart) there is generally a gradient of allele frequency differences among populations consistent with an "isolation by distance" model.[47] At smaller geographical scales, the picture is more complicated, varies geographically, and may be affected by the insecticide application history.

Triatoma dimidiata appears to be a highly mobile species, whether moving among domestic habitats,[48] or between domestic and sylvan habitats[49] with little subdivision among nearby populations. Across large geographical areas and both ecotopes, strong genetic structure was reported.[50,51] In Colombia, genetic differentiation did not strongly correlate with distance (*Isolation by Distance* model), instead subdivision may be more influenced by demographic history, such as the formation of the Isthmus of Panama and the upwelling of the Andes.[52] Less movement is apparent in *T. infestans* populations,[53] for example, following pesticide treatment of houses, nearly all "reinfestants" are survivors or migrants from nearby peridomestic sites,[43,44,54,55] and sylvan *T. infestans* populations are highly structured.[40] Although largely a domestic species, sylvan *R. prolixus* move readily between houses and palm trees.[42] Where sylvan or peridomestic populations are likely to reinfest, an Ecohealth approach to control, such as wall plastering and cement flooring that make the houses refractory to the insects, can provide a community-engaged, cost-effective, environmentally friendly, and long-term approach to vector control[2,56,57] (Fig. 15.9).

3.7 Conclusions and Future Directions

More than 100 years after its discovery, and despite notable successes of intergovernmental control initiatives, Chagas disease remains the most serious of the parasitic diseases affecting Latin America. Substantial challenges remain, such as emergence of insecticide resistance, secondary vectors replacing eliminated primary vectors, and the spread of Chagas as a result of deforestation, climate change, and global travel. Completion of the first Triatominae genome, *R. prolixus*,[58] tissue-specific transcriptomes[59–62] and studies that are underway will provide many new tools to address these challenges. Comparative studies promise advances in understanding the genetic basis of vector competence/capacity, reproductive isolation among sympatric species, as well as genes and proteins involved in the switch to hematophagy, domestication, and insecticide resistance.[63] New hope for eventual elimination of Chagas disease comes from an integrative approach combining new tools in "-omics"

Downloaded From Vet-eBooks.com



Figure 15.9 Photos of the same house before and after the Ecohealth house improvements that make the houses refractory to the triatomine vectors.
Reproduced with permission from C. Monroy.

with mathematical modeling to design evidence-based interventions, in conjunction with community-based development approaches.

4. The *Anopheles gambiae* Complex

4.1 Introduction

The *Anopheles gambiae* species complex was initially described as containing six cryptic (morphologically indistinguishable) species: *A. gambiae sensu strico* Giles,

Downloaded From Vet-eBooks.com

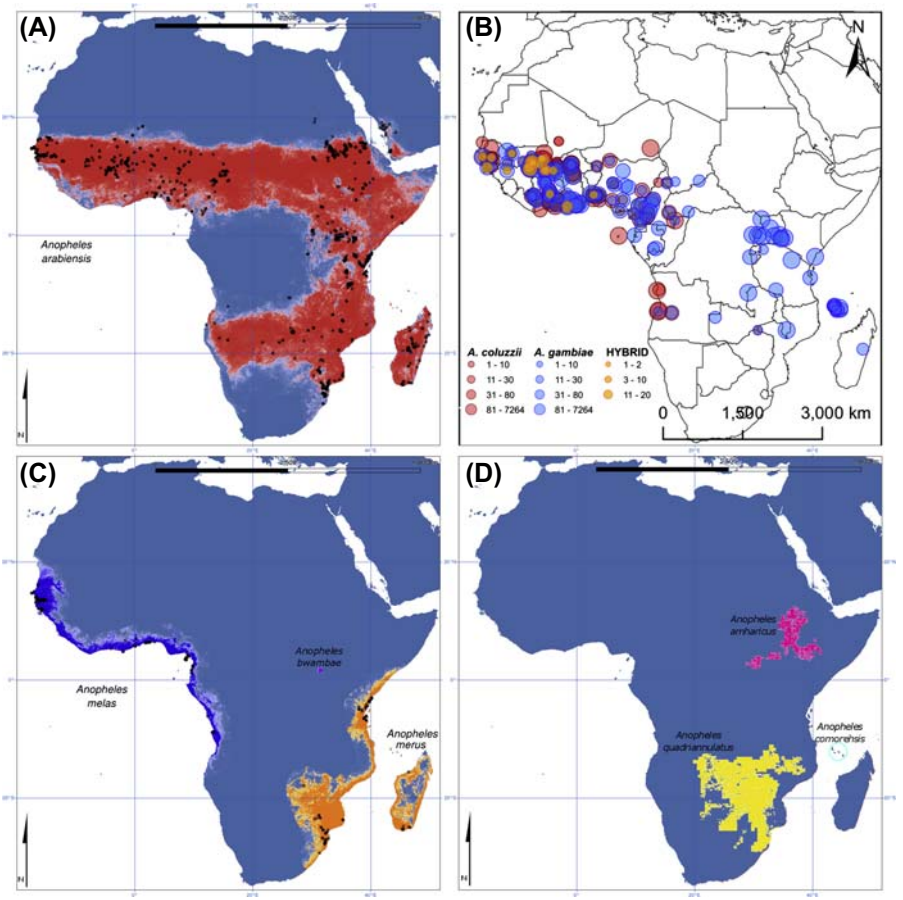


Figure 15.10 Geographical distribution of species in the *Anopheles gambiae* complex (A, C, and D). *(Adapted from Ayala FJ, Coluzzi M. Chromosome speciation: humans, Drosophila, and mosquitoes. Proc Natl Acad Sci USA 2005;**102**(Suppl. 1):6535–42.)* (B) Illustrates the distribution of *A. gambiae* s.s., *A. coluzzii*, and hybrids and provides information of the relative abundance of each.

Anopheles arabiensis Patton, *Anopheles bwambae* White, *Anopheles melas* Theobald, *Anopheles merus* Dönitz, and *Anopheles quadriannulatus* Theobald. Species in the complex are distributed throughout sub-Saharan Africa (Fig. 15.10). The status of these species was initially established via the demonstration of F_1 hybrid sterility among crosses between different *A. gambiae sensu lato* populations.[64–67] Subsequent studies revealed that these six species could be distinguished on the basis of fixed differences in chromosomal inversions.[67–69] Three additional species were later described: *Anopheles comorensis* Brunhes, le Goff & Geoffroy based on subtle morphological features[70]; *Anopheles amharicus* Coetzee, Hunt & Wilkerson[71] based on hybrid male sterility in crosses with *A. quadriannulatis*[72]; and *Anopheles coluzzii* Coetzee, Hunt & Wilkerson, on the basis of an X-linked molecular marker.[73]

Downloaded From Vet-eBooks.com

Of the nine species, three—*A gambiae sensu stricto*, *A. coluzzii*, and *A. arabiensis*—have the broadest geographical distribution (Fig. 15.10) and are the most important vectors of human malaria. *A. coluzzii* and *A. gambiae* s.s. have been the most studied with respect to molecular and population genetics. The whole-genome sequence of *A. coluzzii* was published in 2002.[74] Although the original genome sequence publication was described as being *A. gambiae*, the strain used to generate this sequence (PEST) was an *A. gambiae*/*A.coluzzii* hybrid that had the *A. coluzzii* type X-linked diagnostic sequence, making this technically the *A. coluzzii* genome, not *A. gambiae*.

4.2 *Population Genetic Structure in* Anopheles gambiae

Anopheles gambiae s.l. is structured (i.e., departs from random breeding or panmixia) in at least three ways: (1) temporal—there are seasonal variations in population size and composition; (2) geographical—they mate locally, with little migration among villages; and (3) nondimensional—even within the same location and time, mating is nonrandom.

4.2.1 Temporal Structure

There are seasonal differences in abundance and composition of *A. gambiae* s.l. For example, in Banambani, Mali, *A. arabiensis* and *A. gambiae* s.s. are present in large numbers during the rainy season, with a progressive increase of *A. gambiae* s.s. during the rainy season and *A. arabiensis* in the drier months.[75] Evidence suggests that *A. coluzzii* estivate as adults during the dry season,[76] while *A. gambiae* s.s. is likely to recolonize after local extinction.[77] The pattern varies somewhat from place to place, and is especially different in irrigated areas.[78]

4.2.2 Geographical Structure

The geographical structure is complex and is poorly understood through much of the species range. Gene frequencies at nine microsatellite loci showed that the Rift Valley of East Africa imposes a huge barrier to gene flow among populations of *A. gambiae* s.s.[79] A cluster analysis of a more extensive study based on gene frequencies for 11 microsatellite loci revealed a major subdivision among *A. gambiae* populations in Africa[80]: They identified a northwestern (NW) population group, containing populations in Senegal, Ghana, Nigeria, Cameroon, Gabon, Democratic Republic of Congo, and western Kenya and a southeastern (SE) group including populations in eastern Kenya, Tanzania, Malawi, and Zambia (Fig. 15.11). Differentiation between these two population groups was relatively high ($F_{ST} > 0.1$). A later study[81] corroborated the subdivision between NW and SE groups.

4.2.3 Nondimensional Structure

There is extensive nonrandom mating among genetically distinct subpopulations of *A. gambiae* s.s.[75] and possibly within *A. coluzzii*,[82] known as chromosomal and/or

Downloaded From Vet-eBooks.com

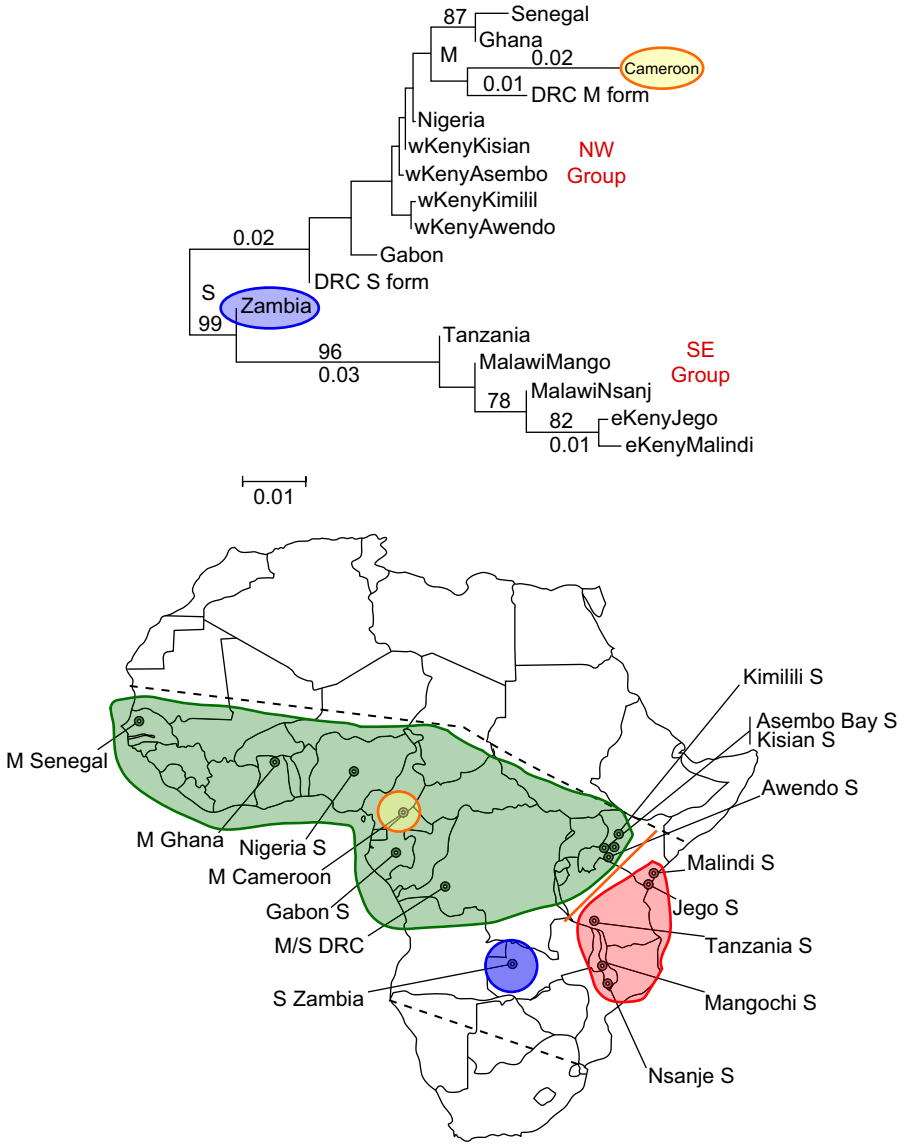


Figure 15.11 Top. Unrooted neighbor−joining population tree based on mean F_{ST} across nine microsatellite loci. M and S populations are denoted at the bases of the clades. The northwestern and southeastern population groups are indicated. Fractions denote branch length (over 0.01) and integers denote biologically significant bootstrap support values. Bottom: Map roughly indicating the boundaries of the different population groups. The orange line separating the southeastern group (in red) from the northwestern group (in green) represents the location of the Great Rift Valley.
Adapted from Lehmann T, Licht M, Elissa N, Maega BT, Chimumbwa JM, Watsenga FT, et al. Population structure of Anopheles gambiae in Africa. *J Hered* 2003;**94**(2):133−47.

Downloaded From Vet-eBooks.com

molecular forms. The amount of gene flow among these populations and between species in Mali has been measured,[83] and gene flow between forms seem internally consistent. However, the amount of hybridization varies considerably from location to location. Because some forms are more persistently present than others, and even absent at some locations, the amount of crossing will vary from place to place. But there are apparently intrinsic factors that also play a role in the degree of between-form hybridization.

4.3 *The Unique Relationship Between* Anopheles coluzzii *and* Anopheles gambiae

Although the distributions of many of the sister species within the *A. gambiae* complex overlap (Fig. 15.10), the occurrence of interspecific hybrids is very rare. One exception is *A. gambiae* and *A. coluzzii* in which hybrids between these two species have been frequently observed (Fig. 15.10B). The unique relationship between *A. coluzzii* and *A. gambiae* merits special consideration.

Anopheles gambiae was long known to exist in nature as two distinct and sympatric populations. Initially described as the M and S molecular forms, these are now recognized as distinct species, designated *A. coluzzii* and *A. gambiae*, respectively.[71] The two exhibit varying degrees of reproductive isolation (RI) across their range.[84–87] Assortative mating is thought to be the major force maintaining RI[88]; however, studies reported in 2013 provide strong evidence for reduced hybrid fitness in nature as an important additional isolating factor.[87]

Divergence in the genomes of *A. gambiae* and *A. coluzzii* is highest at three small regions (~3% of genome) located near the centromeres (pericentromeric), described as islands of speciation, one on each of the three chromosomes—X, 2, and 3.[89,90] These have been described as "islands of speciation" because they are hypothesized to contain genes important in maintaining RI between the two species. This hypothesis has been supported by a 2015 study demonstrating an important role for the X-chromosome island of speciation with respect to mate choice in *A. coluzzii* and *A. gambiae*.[91] A large-scale study utilizing the multiple species-specific SNPs demonstrated that levels of gene flow between *A. coluzzii* and *A. gambiae* are spatially heterogeneous.[87] A longitudinal survey, which was part of this study, revealed temporal heterogeneity as well, with long periods of strong reproductive isolation periodically interrupted by episodes of hybridization (abrupt appearance of F1 genotypes), followed by the disappearance of hybrid genotypes and reestablishment of disequilibrium. A different pattern emerged following an episode of hybridization that occurred in 2006 where SNPs on the X-chromosome and on chromosome 3 returned to disequilibrium, but chromosome 2 SNPs, previously fixed in *A. gambiae*, introgressed into the *A. coluzzii* genome.[92] This genome region was found to contain the knockdown resistance (*kdr*) locus important for resistance to DDT and pyrethroid insecticides, and the resistance allele was present in *A. gambiae*, but absent in *A. coluzzii*. It is thought that in this case, the *kdr* gene was transferred across the species boundary from *A. gambiae* into *A. coluzzii* via a process known as *adaptive introgression*.[92,93]

Downloaded From Vet-eBooks.com

4.4 *Significance of* Anopheles gambiae *Population Genetics to Malaria Transmission and Control*

Ultimately, malaria control efforts in Africa will have to be conducted on a large geographical scale. Sub-Saharan Africa includes a wide variety of ecological zones. It is not surprising that *A. gambiae* s.l., with a distribution across the continent, is highly diverse. The success of malaria control strategies aimed at controlling or manipulating its vectors will have to include knowledge of diversity in vector populations, how this diversity is distributed in time and space, and the forces limiting gene flow and maintaining diversity among populations.

Analyses of the genetic structure of *A. gambiae* s.l. populations have contributed to understanding the distribution of phenotypic variation in the complex and at the subspecific level. The description of *A. quadriannulatis* as a distinct species and recognition that it is primarily zoophyllic provided an explanation for variation in host preference.[94] The evolution of insecticide resistance in populations of *A. coluzzii* poses a serious challenge to current malaria control programs that rely on impregnated bed nets and indoor spraying for vector control.[95,96] The differential distribution of resistance genes, such as *kdr*, among *A. coluzzii* and *A. gambiae* s.s. populations establishes the importance of recognizing population structure to insecticide resistance monitoring.[97–99]

The availability of the *A. gambiae* s.s. whole-genome sequence[74] has ushered in the advent of *population genomics* in vector biology. The promise of establishing the relationship between phenotype and genotype is attainable through the powerful new approach of *association mapping*. Early work aimed at identifying genes directly responsible for phenotypes of interest involved the use of laboratory strains selected for those phenotypes.[100,101] In the studies reported in the late 1990s it was pointed out that this approach has serious limitations and that studies based on natural populations provide far more useful information.[102,103] It is well known that the presence of population structure can result in "spurious associations" between a phenotype and markers that are not linked to any causative loci (e.g., Refs. 104–107). This becomes a problem when these subpopulations are not recognized so that a sample being used in an *association mapping* study consists of a mixture of individuals originating from two or more diverged subpopulations.

The movement of genes, including the use of gene-drive vehicles (e.g., homing endonucleases,[108] clustered regularly interspaced short palindromic repeats (CRISPR)-associated protein 9 (Cas9)-mediated gene-drive[108]) from one lineage or population to another, depends on mating between an individual carrying the gene and one that does not. Although designs for novel approaches to target vector populations of mosquitoes are interesting and potentially useful, the population genetics component is very poorly understood. Critically, most conceptual models for genetic control assume that the mosquito population into which a refractory gene system is to be released represents a single, randomly mating unit. We have summarized the evidence that natural populations of *A. gambiae* are subdivided by barriers to reproduction and that gene flow via migration among geographical populations is limited. Field studies designed to estimate levels and patterns of gene flow within and among natural vector populations are needed to provide a foundation for predicting the potential utility of new molecular-level approaches, and for designing field trials to evaluate their efficacy under natural conditions in Africa.

Downloaded From Vet-eBooks.com

4.5 Conclusions

With respect to current concepts toward describing the genetics of populations of *A. gambiae* s.l.:

- Reproductive isolation among *A. coluzzii* and *A. gambiae* species group appears to be associated with relatively small genomic "islands of divergence" located near the centromere on the X-chromosome and possibly other "islands" located on chromosomes 2 and 3.
- Hybridization between *A. coluzzii* and *A. gambiae* is far more frequent than described in the literature.
- Rates of hybridization between *A. coluzzii* and *A. gambiae* vary both spatially and temporally.

5. Genetics of the Order Ixodida

5.1 Introduction

Ticks (subphylum Chelicerata; subclass Acari; order Ixodida) are global pests affecting human and animal health. Ticks are obligate, blood-feeding ectoparasites recognized for their ability to transmit the broadest spectrum of pathogens including viruses, bacteria, protozoa, fungi, and nematodes to their vertebrate host.[109] Among arthropods, ticks are considered second only to mosquitoes in terms of their importance to public health.[110] The incidence of tick-borne diseases is increasing worldwide and many are considered emerging zoonoses and recognized as potential threats to biosecurity. For reviews, see Refs. 109,111. Diseases transmitted by ticks include Lyme disease (LD), tick-borne relapsing fever (TBRF), babesiosis, anaplasmosis, Rocky Mountain spotted fever (RMSF), Boutonneuse fever, Queensland tick typhus, Q fever, and numerous arboviruses.[112] Also of importance are tick-transmitted zoonoses, such as anaplasmosis, babesiois, theileriosis, and African swine fever that impact livestock production worldwide.[113]

Despite the importance of ticks, little is known about the molecular mechanisms that underpin parasitic processes and pathogen transmission among members of this group. Genetics has been limited by the lack of genetically tractable systems, genetic markers and maps, and transformation tools. Next-generation sequencing (NGS) technologies have enabled major advances in genomic studies of the Acari. This progress has facilitated an improved understanding of tick biology at the molecular level and enabled genetic analyses for many species.

5.2 Systematics, Biogeography, and Medical/Veterinary Significance

The Acari is a diverse group comprising the lineages Acariformes (includes disease-transmitting chigger mites) and Parasitiformes (includes ticks and other medically important mites). The vast majority of species are harmless to humans or beneficial in ecosystems. Approximately 250 species cause health problems for humans and domestic animals.[114] The superorder Parasitiformes includes the order Ixodida

Downloaded From Vet-eBooks.com

comprising the families Ixodidae (hard ticks), Argasidae (soft ticks), and the Nuttallielidae (represented by a single species). There are about 907 valid species of ticks; the majority of species are ectoparasites of wildlife and about 10% are recognized as vectors of disease to humans and animals or for their ability to cause direct damage or paralysis. The family Ixodidae comprises two lineages, the Prostriata consisting of the single genus *Ixodes* (~249 species) and the Metastriata (~464 species, 11 genera) and includes many vectors. The *Ixodes ricinus* species complex, one of the most important affecting public health globally, is comprised of 14 closely related taxa that are distributed in almost all geographical regions of the world.[115]

5.3 Cytogenetics

Most hard ticks studied have an XX−XO (female−male) sex-determination system,[116] whereas sex determination is typically XX−XY in soft ticks.[117] Of the tick species examined,[118] the number of somatic chromosomes ranged from 2 to 36 and the sex chromosome systems included XX−XY, XX−XO, and $X_1X_1X_2X_2$−X_1X_2Y variants. Studies of *Ixodes scapularis* chromosomes revealed $2n = 28$ karyotype and an XX−XY sex-determination system.[119−121] An XX−XO sex-determination system was reported for *Rhipicephalus microplus* (southern cattle tick) with 22 diploid chromosomes in females and 21 in males.[118] The first fluorescence in situ hybridization (FISH)-based karyotype for *R. microplus* was produced in 2009[122] and for *I. scapularis* in 2010.[123] These studies are a step toward a comprehensive understanding of genome organization in pro- and metastriate ticks and the production of integrated physical, genetic, and sequence maps.

5.4 Phylogenetics and Molecular Diagnostics

Numerous studies have analyzed the phylogeny, evolution, and historical zoogeography of the Ixodida.[124] Determination of biosystematic relationships using phenotypic methods has proved difficult. Phylogenetic studies have been limited by the lack of fossil evidence, specimens, and molecular data, and basal relationships within the order remain poorly understood. It has been proposed that the Ixodida evolved from free-living, saprophytic mites[125] and that hard ticks evolved from bird-feeding soft ticks.[126] Several hypotheses have been proposed to explain the origin of the hard ticks.[127,128] The Prostriata are considered a paraphyletic clade with one clade comprising Australasian species and the other, non-Australasian species.[127] The metastriata contains four subfamilies, and several revisions to the phylogeny have been proposed.[129,130] The complete evolutionary history of the Ixodida was described by Mans et al.[131] based on analyses of genes, biochemical systems, morphological characteristics, and phenotypes.

Phylogenetic relationships of the Ixodida have been explored at the genus, family, and subfamily levels using a variety of mitochondrial and nuclear ribosomal DNA (rDNA) markers. For reviews covering the scope of marker-based research on ticks and mites, see Refs. 132−134. Mitochondrial rDNA was used to derive a molecular

Downloaded From Vet-eBooks.com

phylogeny for the Argasidae and Ixodidae that largely supported the work of Hoogstraal and Aeschlimann.[135,136] Phylogenetic relationships of hard and soft ticks at the subfamily level have been explored using nuclear 28S and 18S rDNA[128,137,138] and across the superorder Parasitiformes.[139] Variation in the internal transcribed spacer (ITS) has been employed to distinguish species in multiple genera.[140–142] Biochemical and molecular techniques are available to identify ticks[143] and DNA barcoding capabilities are under development for the Ixodidae,[144] but molecular diagnostic tools remain a significant need for multiple species.

5.5 Genetic Diversity and Population Genetics

Studies of genetic diversity have been reported for at least 22 tick species from six genera, representing the Argasidae and Ixodidae.[134] Observed levels of population genetic structure range from negligible to high across the Ixodida, and for some species, suggest a correlation to host movement and significant host-race adaptation.[134] Polymorphic microsatellite loci have been identified in *I. scapularis*, *I. ricinus*, and *Ixodes uriae*[145,146] and may prove useful for resolving genetic variation at the inter- and intraspecies level. Thousands of single nucleotide polymorphism (SNP) markers have enabled detailed analyses of genetic diversity in *I. scapularis* populations[147–149] and these markers will have broad utility for studies across the Acari.

The evolutionary history of *I. scapularis* has been inferred based on mitochondrial 16S and other DNA markers.[150] Studies indicate *I. scapularis* was restricted to southern North America during Pleistocene glaciation events with recolonization of northern North America by founding populations after the recession of ice sheets.[148,151] Two distinct clades are recognized that exhibit behavioral and morphological variations. The "All American Clade" occurs in northern and southern states and the more genetically diverse "Southern Clade" is found only in the southern United States.[152] Analyses involving hundreds of SNP markers revealed signatures for migration of northern ticks into southern populations[148] and raised concerns given the greater ability of northern *I. scapularis* to vector *Borrelia burgdorferi*. An extensive analysis of genetic diversity was conducted among eight populations of *I. scapularis* from the northeast, mid-west, and southeast regions of the United States using the restriction site–associated DNA sequencing (RADseq) technique.[149] Results suggest low levels of inbreeding between populations and support a single species classification across North America as previously proposed.[119] Genome-wide analyses of population structure revealed five clades and signatures of north–south structure (Fig. 15.12). Results support a genetic component associated with differences in the natural history of *I. scapularis* populations and a correlation to the prevalence of human LD cases, and highlight opportunities to identify loci associated with pathogen transmission by the tick.

5.6 Genomics and Genetic Mapping

Among the Acari, progress in genome sequencing has been greatest for species of mites. Assembled and annotated genomes are available for the two-spotted spider

Downloaded From Vet-eBooks.com

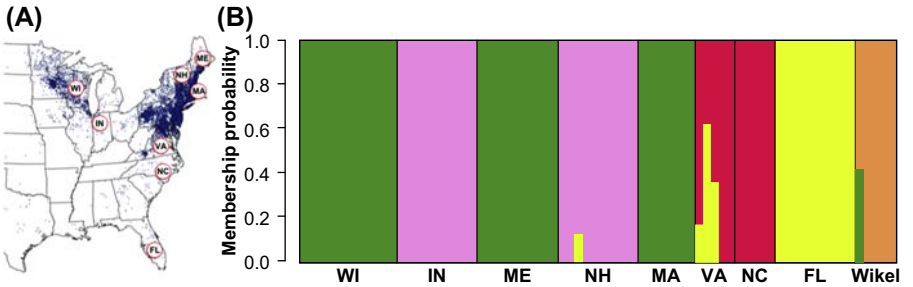


Figure 15.12 Population structure of *Ixodes scapularis* across North America. (A) Tick sampling sites in Indiana (IN), Massachusetts (MA), Maine (ME), North Carolina (NC), New Hampshire (NH), Wisconsin (WI), Florida (FL), and Virginia (VA) overlaid against reported Lyme disease cases in 2012. *(Modified from CDC: http://www.cdc.gov/lyme/stats/maps/map2012.html.)* (B) Membership probabilities in bar plots for individual *I. scapularis* comprising different clusters and showing separation of genetic groups based on 34,693 RADtag SNP markers. *SNP*, single nucleotide polymorphism; *WK, I. scapularis*, WIKEL reference strain. *(Image reprinted by permission from Gulia-Nuss M, Nuss AB, Meyer JM, Sonenshine DE, Roe RM, Waterhouse RM, et al. Genomic insights into the Ixodes scapularis tick vector of Lyme disease. Nature communications 2016;7:10507.* http://dx.doi.org/10.1038/ncomms10507. *Macmillan Publishers Ltd., copyright 2016.)* The reader is referred to the online version for color.

mite *Tetranychus urticae*,[153] honey bee parasite *Varroa destructor*,[154] house dust mite *Dermatophagoides farinae*,[155] predatory mite *Metaseiulus occidentalis*,[156,157] and scabies mite *Sarcoptes scabiei* var. *canis*.[158] The assembly and annotation of the about 2.1 Gbp *I. scapularis* genome[149] is the first such project for a tick vector (ABJB010000000). Approximately 20% of *I. scapularis* gene models are unique to the tick and are a source of potential novel targets for discovery of new vaccines and insecticides. Sequencing of the about 7.1 Gbp *R. microplus* genome has also been proposed,[159] making this tick the representative species for genomic research within the metastriate lineage.[160]

In the postgenomic era, research has been aimed at understanding genome evolution and composition at finer scale. Several large-scale gene duplication events in *I. scapularis* may be critical to the success of ticks as parasites.[161] Progress in gene discovery for pro- and metastriate ticks, as well as soft ticks, has been extensive[133] with an emphasis on elucidating gene products associated with tick−host−pathogen interactions.[162−164] The preliminary *I. scapularis* linkage map[146] represents an important tool to identify genes associated with host preference, vector competence, and acaricide resistance and is the only such resource available for any tick in the order Ixodida. Efforts are ongoing to develop a high-density linkage map based on thousands of SNP markers reported by Gulia-Nuss et al.[149] (Hill *pers. comm.*). The development of genetic markers and mapping populations for other species of ixodid ticks, particularly *R. microplus*, is a priority and will support mapping of quantitative trait loci (QTLs) and the assembly of physical- and sequence-based maps.[133]

Downloaded From Vet-eBooks.com

5.7 Perspectives for Control and Future Directions

As of 2016, the notable accomplishments in the field of tick genetics and genomics in the last several decades include the development of molecular phylogenies, studies of genetic population structure and descriptions of the transcriptome and proteome of important vectors, functional studies of multiple gene products, and more recently, whole-genome sequencing and population genomics.[133,149] We can anticipate the field embracing genetic control strategies (e.g., CRISPR/Cas9 gene editing) for manipulating populations of major pests. Key to these efforts will be studies to pinpoint loci associated with phenotypes of interest. Also on the horizon is the translation of genetic data into highly specific vaccine and acaricide products for disease control.[149] The development of genetic resources such as phenotypic and molecular markers, physical, sequence, and linkage maps, genetic lines and mapping populations, in vitro and in vivo transformation capabilities, and tools to model and manage control strategies within population frameworks will be essential for tick and tick-borne disease control.

Glossary

Allopatric taxa are occupying distinctly different geographical ranges.

Association mapping is a method of gene mapping that utilizes historic linkage disequilibrium (linkage) to associate phenotypes to genotypes.

Cryptic species are species that are identical or nearly identical in appearance, but can be discovered by genetic divergence as indicated by mating incompatibilities or sterile matings in interspecific crosses.

Cytoplasmic incompatibility refers to sterile matings between *Wolbachia*-infected males and uninfected females. Females carrying reproductive parasite *Wolbachia* are fertile when mated with uninfected or *Wolbachia*-infected flies. In this way the maternal host lineages with *Wolbachia* can displace uninfected, maternal lineages if their Darwinian fitnesses are adequate.

Effective population size (N_e) and gene flow N_e is the hypothetical number of reproducing organisms in an ideal population (i.e., obeying Hardy−Weinberg assumptions) that corresponds to the population under investigation. Gene flow is the number of effective migrants N_em exchanged per generation among subpopulations. $N_em = (1-F_{ST})/4F_{ST}$ for diploid loci.

F_{ST} and F_{IS} are inbreeding coefficients. F_{ST} measures departures from random mating among subpopulations and is inversely related to gene flow. F_{IS} measures departures from random mating within subpopulations.

Incipient species populations evolving toward complete reproductive isolation and therefore distinct biological species status.

Introgression is introduction of genetic material from another species or variant into a population by hybridization followed by repeated backcrossing.

Downloaded From Vet-eBooks.com

Isolation by distance the probability of mating between two individuals decreases with increasing distance between them, resulting in a direct relationship between geographical and genetic distance.

Molecular forms are populations of *A. gambiae* s.s. that differ with respect to specific sequence in a region of the intergenic spacer segment of the ribosomal gene, as visualized using diagnostic PCR methods.

Monophyletic taxon is a group of organisms including the most recent common ancestor of all those organisms and all the descendants of that common ancestor.

Paracentric chromosome inversion of a segment of a chromosome does not include the centromere. A *pericentric inversion* does include the centromere.

Peridomestic habitats are the area surrounding houses, for example, wood piles, animal corrals, and so on.

Polyphyletic taxon is a group of organisms derived from two or more ancestral lineages.

Polytypic species is a species that contains several variant forms, especially geographically or temporally differentiated subspecies or varieties, which would normally interbreed if present in the same time and place.

Population genomics sampling of numerous, or all, variable gene loci within a genome to infer those evolutionary forces responsible for observed patterns of variation.

Acknowledgments

The authors would like to gratefully acknowledge the contributions to a previous version by Francois Noireau and technical assistance of Adrienne Woods. This work was made possible in part by support from National Science Foundation (NSF) grant BCS-1216193 as part of the joint NSF-NIH-USDA (United States Department of Agriculture) Ecology and Evolution of Infectious Diseases program.

References

1. World Health Organization. *Vector-borne diseases, Fact sheet no. 387*. 2014. Available from: http://www.who.int/mediacentre/factsheets/fs387/en/.
2. Monroy C, Castro X, Bustamante DM, Pineda SF, Rodas A, Moguel B, et al. An ecosystem approach for the prevention of Chagas disease in rural Guatemala. In: Charron DF, editor. *Ecohealth research in practice: innovative applications of an ecosystem approach to health*. New York: Springer; 2012.
3. Krafsur ES. The sterile insect technique. In: Pimentel D, editor. *Encyclopedia of pest management*. London: Taylor & Francis; 2002. p. 788–91.
4. Simarro PP, Cecchi G, Franco JR, Paone M, Diarra A, Ruiz-Postigo JA, et al. Estimating and mapping the population at risk of sleeping sickness. *PLoS Neglected Trop Dis* 2012; **6**(10):e1859.
5. Reinhardt E. *Travailler ensemble: la mouche tsé-tsé et la pauvreté rurale*. 2002. Available from: http://www.un.org/french/pubs/chronique/2002/numero2/0202p17_la_mouche_tsetse.html.

Downloaded From Vet-eBooks.com

6. Aksoy S, Caccone A, Galvani AP, Okedi LM. *Glossina fuscipes* populations provide insights for human African trypanosomiasis transmission in Uganda. *Trends Parasitol* 2013;**29**(8):394−406.
7. Krafsur ES. Tsetse flies: genetics, evolution, and role as vectors. *Infect Genet Evol* 2009; **9**(1):124−41.
8. Doudoumis V, Alam U, Aksoy E, Abd-Alla AM, Tsiamis G, Brelsfoard C, et al. Tsetse-Wolbachia symbiosis: comes of age and has great potential for pest and disease control. *J Invertebr Pathol* 2013;**112**(Suppl:S94−103).
9. Gooding RH, Krafsur ES. Tsetse genetics: contributions to biology, systematics, and control of tsetse flies. *Annu Rev Entomol* 2005;**50**:101−23.
10. Dyer NA, Furtado A, Cano J, Ferreira F, Odete Afonso M, Ndong-Mabale N, et al. Evidence for a discrete evolutionary lineage within Equatorial Guinea suggests that the tsetse fly *Glossina palpalis palpalis* exists as a species complex. *Mol Ecol* 2009;**18**(15): 3268−82.
11. Cordon-Obras C, Cano J, Knapp J, Nebreda P, Ndong-Mabale N, Ncogo-Ada PR, et al. *Glossina palpalis palpalis* populations from Equatorial Guinea belong to distinct allopatric clades. *Parasit Vectors* 2014;**7**:31.
12. De Meeûs T, Bouyer J, Ravel S, Solano P. Ecotype evolution in *Glossina palpalis* subspecies, major vectors of sleeping sickness. *PLoS Neglected Trop Dis* 2015;**9**(3): e0003497.
13. Gooding R. Genetic analysis of sterility in hybrids from crosses of *Glossina morsitans submorsitans* and *Glossina morsitans centralis* (Diptera: Glossinidae). *Can J Zool* 1993; **71**:963−72.
14. Machado AB. *Révision systématique des glossines du groupe palpalis (Diptera)*. Lisboa: Museo do Dundo, Companhia de Diamantes de Angola; 1954. 189 pp.
15. Dyer NA, Ravel S, Choi KS, Darby AC, Causse S, Kapitano B, et al. Cryptic diversity within the major trypanosomiasis vector *Glossina fuscipes* revealed by molecular markers. *PLoS Neglected Trop Dis* 2011;**5**(8):e1266.
16. Ford J. *The role of Trypanosomiases in African ecology*. Oxford: Clarendon Press; 1971.
17. Rogers DJ. Study of a natural population of tsetse flies and a model for fly movement. *J Animal Ecol* 1977;**46**:309−30.
18. Leak SGA. *Tsetse biology and ecology: their role in the epidemiology and control of trypanosomiasis*. New York: CABI Publ; 1998.
19. Hargrove JW. *Tsetse eradication: sufficiency, necessity and desirability*. Edinburgh: DFID Animal Health Programme, Centre for Tropical Veterinary Medicine, University of Edinburgh; 2003. UK.
20. Rogers DJ, Robinson TP. Tsetse distribution. In: Maudlin I, Holmes P, Miles M, editors. *The Trypanosomiases*. Oxford, UK: CABI; 2004. p. 139−79.
21. Bourtzis K, Lees RS, Hendrichs J, Vreysen MJ. More than one rabbit out of the hat: radiation, transgenic and symbiont-based approaches for sustainable management of mosquito and tsetse fly populations. *Acta Trop* 2016;**157**:115−30.
22. Alam U, Medlock J, Brelsfoard C, Pais R, Lohs C, Balmand S, et al. *Wolbachia* symbiont infections induce strong cytoplasmic incompatibility in the tsetse fly *Glossina morsitans*. *PLoS Pathog* 2011;**7**(12):e1002415.
23. Vale GA, Torr SJ. User-friendly models of the costs and efficacy of tsetse control: application to sterilizing and insecticidal techniques. *Med Vet Entomol* 2005;**19**(3): 293−305.
24. Torr SJ, Hargrove JW, Vale GA. Towards a rational policy for dealing with tsetse. *Trends Parasitol* 2005;**21**(11):537−41.

Downloaded From Vet-eBooks.com

25. *Weekly epidemiology report: Chagas disease in Latin America: an epidemiological update based on 2010 estimates*. Switzerland: World Health Organization; 2015.
26. Weirauch C, Munro JB. Molecular phylogeny of the assassin bugs (Hemiptera: Reduviidae), based on mitochondrial and nuclear ribosomal genes. *Mol Phylogenet Evol* 2009; **53**(1):287−99.
27. Patterson JS, Gaunt MW. Phylogenetic multi-locus codon models and molecular clocks reveal the monophyly of haematophagous reduviid bugs and their evolution at the formation of South America. *Mol Phylogenet Evol* 2010;**56**(2):608−21.
28. Justi SA, Galvão C, Schrago CG. Geological changes of the Americas and their influence on the diversification of the Neotropical kissing bugs (Hemiptera: Reduviidae: Triatominae). *PLoS Neglected Trop Dis* 2016.
29. Paula AS, Diotaiuti L, Schofield CJ. Testing the sister-group relationship of the Rhodniini and Triatomini (Insecta: Hemiptera: Reduviidae: Triatominae). *Mol Phylogenet Evol* 2005; **35**(3):712−8.
30. Hwang WS, Weirauch C. Evolutionary history of assassin bugs (insecta: hemiptera: Reduviidae): insights from divergence dating and ancestral state reconstruction. *PLoS One* 2012;**7**(9):e45523.
31. Justi SA, Russo CA, Mallet JR, Obara MT, Galvao C. Molecular phylogeny of Triatomini (Hemiptera: Reduviidae: Triatominae). *Parasit Vectors* 2014;**7**:149.
32. Lent H, Wygodzinsky P. Revision of the Triatominae (Hemiptera, Reduviidae) and their significance as vectors of Chagas disease. *Bull Am Mus Nat Hist* 1979;**163**:123−520.
33. Ibarra-Cerdena CN, Zaldivar-Riveron A, Peterson AT, Sanchez-Cordero V, Ramsey JM. Phylogeny and niche conservatism in North and Central American triatomine bugs (Hemiptera: Reduviidae: Triatominae), vectors of chagas' disease. *PLoS Neglected Trop Dis* 2014;**8**(10):e3266.
34. Hypsa V, Tietz DF, Zrzavy J, Rego RO, Galvao C, Jurberg J. Phylogeny and biogeography of Triatominae (Hemiptera: Reduviidae): molecular evidence of a new world origin of the Asiatic clade. *Mol Phylogenet Evol* 2002;**23**(3):447−57.
35. Neiva A, Pinto C. Dos reduvideos hematophagos encontrados no Distrito Federal e Estado do Rio de Janeiro, com a descripsao de uma especie nova. *Brasil-Medico* 1923;**37**:45−7.
36. Hashimoto K, Schofield CJ. Elimination of *Rhodnius prolixus* in Central America. *Parasit Vectors* 2012;**5**:45.
37. Panzera F, Perez R, Panzera Y, Ferrandis I, Ferreiro MJ, Calleros L. Cytogenetics and genome evolution in the subfamily Triatominae (Hemiptera, Reduviidae). *Cytogenet Genome Res* 2010;**128**:77−87.
38. Panzera F, Ferrandis I, Ramsey J, Ordonez R, Salazar-Schettino PM, Cabrera M, et al. Chromosomal variation and genome size support existence of cryptic species of *Triatoma dimidiata* with different epidemiological importance as Chagas disease vectors. *Trop Med Int Health* 2006;**11**(7):1092−103.
39. Schofield CJ, Diotaiuti L, Dujardin JP. The process of domestication in Triatominae. *Mem Inst Oswaldo Cruz* 1999;**94**(Suppl. 1):375−8.
40. Waleckx E, Salas R, Huaman N, Buitrago R, Bosseno MF, Aliaga C, et al. New insights on the Chagas disease main vector *Triatoma infestans* (Reduviidae, Triatominae) brought by the genetic analysis of Bolivian sylvatic populations. Infection, genetics and evolution. *J Mol Epidemiol Evol Genet Infect Dis* 2011;**11**(5):1045−57.
41. Piccinali RV, Marcet PL, Noireau F, Kitron U, Gürtler RE, Dotson EM. Molecular population genetics and phylogeography of the Chagas disease vector *Triatoma infestans* in South America. *J Med Entomol* 2009;**46**(4):796−809.

Downloaded From Vet-eBooks.com

42. Fitzpatrick S, Feliciangeli MD, Sanchez-Martin MJ, Monteiro FA, Miles MA. Molecular genetics reveal that silvatic *Rhodnius prolixus* do colonise rural houses. *PLoS Neglected Trop Dis* 2008;**2**(4):e210.
43. Garcia BA, Manfredi C, Fichera L, Segura EL. Short report: variation in mitochondrial 12S and 16S ribosomal DNA sequences in natural populations of *Triatoma infestans* (Hemiptera: Reduviidae). *Am J Trop Med Hyg* 2003;**68**(6):692−4.
44. Pérez de Rosas AR, Segura EL, García BA. Microsatellite analysis of genetic structure in natural *Triatoma infestans* (Hemiptera: Reduviidae) populations from Argentina: its implication in assessing the effectiveness of Chagas' disease vector control programmes. *Mol Ecol* 2007;**16**(7):1401−12.
45. Pérez de Rosas AR, Segura EL, Fichera L, Garcia BA. Macrogeographic and microgeographic genetic structure of the Chagas' disease vector *Triatoma infestans* (Hemiptera: Reduviidae) from Catamarca, Argentina. *Genetica* 2008;**133**(3):247−60.
46. Schweigmann N, Vallve S, Muscio O, Ghillini M, Alberti A, Wisnivesky-Colli C. Dispersal flight by *Triatoma infestans* in an arid area of Argentina. *Med Vet Entomol* 1988; **2**(4):401−4.
47. Wright S. Isolation by distance. *Genetics* 1943;**28**:114−38.
48. Stevens L, Monroy MC, Rodas AG, Hicks RM, Lucero DE, Lyons LA, et al. Migration and gene flow among domestic populations of the Chagas insect vector *Triatoma dimidiata* (Hemiptera: Reduviidae) detected by microsatellite loci. *J Med Entomol* 2015;**52**(3): 419−28.
49. Dumonteil E, Tripet F, Ramirez-Sierra MJ, Payet V, Lanzaro G, Menu F. Assessment of *Triatoma dimidiata* dispersal in the Yucatan Peninsula of Mexico by morphometry and microsatellite markers. *Am J Trop Med Hyg* 2007;**76**(5):930−7.
50. Blandon-Naranjo M, Zuriaga MA, Azofeifa G, Zeledon R, Bargues MD. Molecular evidence of intraspecific variability in different habitat-related populations of *Triatoma dimidiata* (Hemiptera: Reduviidae) from Costa Rica. *Parasitol Res* 2010;**106**(4):895−905.
51. Gomez-Palacio A, Triana O, Jaramillo ON, Dotson EM, Marcet PL. Eco-geographical differentiation among Colombian populations of the Chagas disease vector *Triatoma dimidiata* (Hemiptera: Reduviidae). *Infect Genet Evol* 2013;**20**:352−61.
52. Gomez-Palacio A, Triana O. Molecular evidence of demographic expansion of the Chagas disease vector *Triatoma dimidiata* (Hemiptera, Reduviidae, Triatominae) in Colombia. *PLoS Neglected Trop Dis* 2014;**8**(3):e2734.
53. Brenière SF, Bosseno MF, Vargas F, Yaksic N, Noireau F, Noel S, et al. Smallness of the panmictic unit of *Triatoma infestans* (Hemiptera: Reduviidae). *J Med Entomol* 1998;**35**: 911−7.
54. Dujardin JP, Cardozo L, Schofield C. Genetic analysis of *Triatoma infestans* following insecticidal control interventions in central Bolivia. *Acta Trop* 1996;**61**(3):263−6.
55. Pizarro JC, Gilligan LM, Stevens L. Microsatellites reveal a high population structure in *Triatoma infestans* from Chuquisaca, Bolivia. *PLoS Neglected Trop Dis* 2008;**2**(3):e202.
56. Feliciangeli MD, Campbell-Lendrum D, Martinez C, Gonzalez D, Coleman P, Davies C. Chagas disease control in Venezuela: lessons for the Andean region and beyond. *Trends Parasitol* 2003;**19**(1):44−9.
57. Waleckx E, Camara-Mejia J, Ramirez-Sierra MJ, Cruz-Chan V, Rosado-Vallado M, Vazquez-Narvaez S, et al. An innovative ecohealth intervention for Chagas disease vector control in Yucatan, Mexico. *Trans R Soc Trop Med Hyg* 2015;**109**(2):143−9.
58. Mesquita RD, Vionette-Amaral RJ, Lowenberger C, Rivera-Pomar R, Monteiro FA, Minx P, et al. Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals

Downloaded From Vet-eBooks.com

unique adaptations to hematophagy and parasite infection. *Proc Natl Acad Sci USA* 2015; **112**(48):14936−41.
59. Ribeiro JMC, Assumpcao TC, Francischetti CN. An insight into the sialomes of blood-sucking Heteroptera. *Psyche* 2012;**2012**:1−16.
60. Schwarz A, Medrano-Mercado N, Schaub GA, Struchiner CJ, Bargues MD, Levy MZ, et al. An updated insight into the Sialotranscriptome of *Triatoma infestans*: developmental stage and geographic variations. *PLoS Neglected Trop Dis* 2014;**8**(12):15.
61. Ons S, Lavore A, Sterkel M, Wulff JP, Sierra I, Martinez-Barnetche J, et al. Identification of G protein coupled receptors for opsines and neurohormones in *Rhodnius prolixus*. Genomic and transcriptomic analysis. *Insect Biochem Mol Biol* 2016;**69**:34−50.
62. Ribeiro JM, Genta FA, Sorgine MH, Logullo R, Mesquita RD, Paiva-Silva GO, et al. An insight into the transcriptome of the digestive tract of the bloodsucking bug, *Rhodnius prolixus*. *PLoS Neglected Trop Dis* 2014;**8**(1):e2594.
63. Mougabure-Cueto G, Picollo MI. Insecticide resistance in vector Chagas disease: evolution, mechanisms and management. *Acta Trop* 2015;**149**:70−85.
64. Davidson G. Insecticide resistance in *Anopheles gambiae* Giles: a case of simple mendelian inheritance. *Nature* 1956;**178**(4538):863−4.
65. Davidson G. The five mating-types in the *Anopheles gambiae* complex. *Riv Malariol* 1964;**43**:167−83.
66. Davidson G. *Anopheles gambiae*, a complex of species. *Bull World Health Organ* 1964; **31**:625−34.
67. Davidson G, Hunt RH. The crossing and chromosome characteristics of a new, sixth species in the *Anopheles gambiae* complex. *Parassitologia* 1973;**15**(1):121−8.
68. Davidson G, Paterson HE, Coluzzi M, Mason GF, Micks DW. The *Anopheles gambiae* complex. *Genet Insect Vectors Dis* 1967.
69. Coluzzi M, Sabatini A, Petrarca V, Di Deco MA. Chromosomal differentiation and adaptation to human environments in the *Anopheles gambiae* complex. *Trans R Soc Trop Med Hyg* 1979;**73**(5):483−97.
70. Brunhes J, Le Goff G, Geoffroy B. *Anophèles* Afro-Tropicaux. I.- Desciptions D'Espèces nouvelles et changements de statuts taxonomiques (Diptera:Culicidae). *Ann Soc Entomol Fr (N S)* 1997;**33**:173−83.
71. Coetzee M, Hunt RH, Wilkerson R, Della Torre A, Coulibaly MB, Besansky NJ. *Anopheles coluzzii* and *Anopheles amharicus*, new members of the *Anopheles gambiae* complex. *Zootaxa* 2013;**3619**:246−74.
72. Hunt RH, Coetzee M, Fettene M. The *Anopheles gambiae* complex: a new species from Ethiopia. *Trans R Soc Trop Med Hyg* 1998;**92**(2):231−5.
73. Fanello C, Santolamazza F, della Torre A. Simultaneous identification of species and molecular forms of the *Anopheles gambiae* complex by PCR-RFLP. *Med Vet Entomol* 2002;**16**(4):461−4.
74. Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, et al. The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 2002;**298**(5591): 129−49.
75. Touré YT, Petrarca V, Traoré SF, Coulibaly A, Maiga HM, Sankaré O, et al. The distribution and inversion polymorphism of chromosomally recognized taxa of the *Anopheles gambiae* complex in Mali, West Africa. *Parassitologia* 1998;**40**(4):477−511.
76. Lehmann T, Dao A, Yaro AS, Adamou A, Kassogue Y, Diallo M, et al. Aestivation of the African malaria mosquito, *Anopheles gambiae* in the Sahel. *Am J Trop Med Hyg* 2010; **83**(3):601−6.

Downloaded From Vet-eBooks.com

77. Mamai W, Simard F, Couret D, Ouedraogo GA, Renault D, Dabire KR, et al. Monitoring dry season persistence of *Anopheles gambiae* s.l. Populations in a contained semi-field system in southwestern Burkina Faso. *West Afr J Med Entomol* 2015.
78. Diuk-Wasser MA, Dolo G, Bagayoko M, Sogoba N, Toure MB, Moghaddam M, et al. Patterns of irrigated rice growth and malaria vector breeding in Mali using multi-temporal ERS-2 synthetic aperture radar. *Int J Remote Sens* 2006;**27**(3):535−48.
79. Lehmann T, Hawley WA, Grebert H, Danga M, Atieli F, Collins FH. The Rift Valley complex as a barrier to gene flow for *Anopheles gambiae* in Kenya. *J Hered* 1999;**90**(6): 613−21.
80. Lehmann T, Licht M, Elissa N, Maega BT, Chimumbwa JM, Watsenga FT, et al. Population structure of *Anopheles gambiae* in Africa. *J Hered* 2003;**94**(2):133−47.
81. Wang-Sattler R, Blandin S, Ning Y, Blass C, Dolo G, Touré YT, et al. Mosaic genome architecture of the *Anopheles gambiae* species complex. *PLoS One* 2007;**2**(11):e1249.
82. Lee Y, Cornel AJ, Meneses CR, Fofana A, Andrianarivo AG, McAbee RD, et al. Ecological and genetic relationships of the Forest-M form among chromosomal and molecular forms of the malaria vector *Anopheles gambiae* sensu stricto. *Malar J* 2009;**8**:75.
83. Taylor C, Toure YT, Carnahan J, Norris DE, Dolo G, Traore SF, et al. Gene flow among populations of the malaria vector, *Anopheles gambiae*, in Mali, West Africa. *Genetics* 2001;**157**(2):743−50.
84. della Torre A, Fanello C, Akogbeto M. Molecular evidence of incipient speciation within *Anopheles gambiae* s.s. in West Africa. *Insect Mol Biol* 2001;**10**:9−18.
85. Marsden CD, Lee Y, Nieman CC, Sanford MR, Dinis J, Martins C, et al. Asymmetric introgression between the M and S forms of the malaria vector, *Anopheles gambiae*, maintains divergence despite extensive hybridization. *Mol Ecol* 2011;**20**(23):4983−94.
86. Riehle MM, Guelbeogo WM, Gneme A, Eiglmeier K, Holm I, Bischoff E, et al. A cryptic subgroup of *Anopheles gambiae* is highly susceptible to human malaria parasites. *Science* 2011;**331**(6017):596−8.
87. Lee Y, Marsden CD, Norris LC, Collier TC, Main BJ, Fofana A, et al. Spatiotemporal dynamics of gene flow and hybrid fitness between the M and S forms of the malaria mosquito, *Anopheles gambiae*. *Proc Natl Acad Sci USA* 2013;**110**(49):19854−9.
88. Tripet F, Toure YT, Taylor CE, Norris DE, Dolo G, Lanzaro GC. DNA analysis of transferred sperm reveals significant levels of gene flow between molecular forms of *Anopheles gambiae*. *Mol Ecol* 2001;**10**(7):1725−32.
89. Turner TL, Hahn MW, Nuzhdin SV. Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol* 2005;**3**(9):e285.
90. White BJ, Cheng C, Simard F, Costantini C, Besansky NJ. Genetic association of physically unlinked islands of genomic divergence in incipient species of *Anopheles gambiae*. *Mol Ecol* 2010;**19**(5):925−39.
91. Aboagye-Antwi F, Alhafez N, Weedall GD, Brothwood J, Kandola S, Paton D, et al. Experimental swap of *Anopheles gambiae's* assortative mating preferences demonstrates key role of X-chromosome divergence island in incipient sympatric speciation. *PLoS Genet* 2015;**11**(4):e1005141.
92. Norris LC, Main BJ, Lee Y, Collier TC, Fofana A, Cornel AJ, et al. Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. *Proc Natl Acad Sci USA* 2015;**112**(3):815−20.
93. Clarkson CS, Weetman D, Essandoh J, Yawson AE, Maslen G, Manske M, et al. Adaptive introgression between *Anopheles* sibling species eliminates a major genomic island but not reproductive isolation. *Nat Commun* 2014;**5**:4248.

Downloaded From Vet-eBooks.com

94. Gillies MT, De Meillon B. The Anophelinae of Africa south of the Sahara (Ethiopian Zoogeographical region). *Publ S Afr Inst Med Res* 1968;**54**:1−343.
95. N'Guessan R, Boko P, Odjo A, Chabi J, Akogbeto M, Rowland M. Control of pyrethroid and DDT-resistant *Anopheles gambiae* by application of indoor residual spraying or mosquito nets treated with a long-lasting organophosphate insecticide, chlorpyrifos-methyl. *Malar J* 2010;**9**:44.
96. Mitri C, Markianos K, Guelbeogo WM, Bischoff E, Gneme A, Eiglmeier K, et al. The kdr-bearing haplotype and susceptibility to *Plasmodium falciparum* in *Anopheles gambiae*: genetic correlation and functional testing. *Malar J* 2015;**14**(1):391.
97. Chandre F, Manguin S, Brengues C, Dossou Yovo J, Darriet F, Diabate A, et al. Current distribution of a pyrethroid resistance gene (*kdr*) in *Anopheles gambiae* complex from west Africa and further evidence for reproductive isolation of the Mopti form. *Parassitologia* 1999;**41**(1−3):319−22.
98. Weill M, Chandre F, Brengues C, Manguin S, Akogbeto M, Pasteur N, et al. The *kdr* mutation occurs in the Mopti form of *Anopheles gambiae* s.s. through introgression. *Insect Mol Biol* 2000;**9**(5):451−5.
99. Tripet F, Wright J, Cornel A, Fofana A, McAbee R, Meneses C, et al. Longitudinal survey of knockdown resistance to pyrethroid *(kdr)* in Mali, West Africa, and evidence of its emergence in the Bamako form of *Anopheles gambiae* s.s. *Am J Trop Med Hyg* 2007; **76**(1):81−7.
100. Sinden EE. Mosquito-malaria interactions: a reappraisal of the concepts of susceptibility and refractoriness. *Insect Biochem Mol Biol* 2004;**34**:625−9.
101. Aguilar R, Dong Y, Warr E, Dimopoulos G. *Anopheles* infection responses; laboratory models versus field malaria transmission systems. *Acta Trop* 2005;**95**(3):285−91.
102. Tripet F, Aboagye'antwi F, Hurd H. Ecological immunology of mosquito-malaria interactions. *Trends Parasitol* 2008;**24**:219−27.
103. Boëte C. *Anopheles* mosquitoes: not just flying malaria vectors... especially in the field. *Trends Parasitol* 2009;**25**(2):53−5.
104. Lander ES, Schork NJ. Genetic dissection of complex traits. *Science* 1994;**265**(5181): 2037−48.
105. Ewens WJ, Spielman RS. The transmission/disequilibrium test: history, subdivision, and admixture. *Am J Hum Genet* 1995;**57**(2):455−64.
106. Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* 1999;**65**(1):220−8.
107. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, et al. Efficient control of population structure in model organism association mapping. *Genetics* 2008; **178**(3):1709−23.
108. Gantz VM, Bier E. Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science* 2015;**348**(6233):442−4.
109. Jongejan F, Uilenberg G. The global importance of ticks. *Parasitol* 2004;**129**:S3−14.
110. Fivaz B, Petney T, Horak I. *Tick vector biology: medical and veterinary aspects*. Heidelberg, Germany: Springer-Verlag; 1992.
111. Nicholson W, Sonenshine D, Lane R, Uilenberg G. In: Mullen GR, Durden LA, editors. *Ticks (Ixodida)*. San Diego, CA: Academic Press; 2009.
112. Sonenshine D. *Biology of ticks*. New York, NY, USA: Oxford University Press; 1991.
113. Hajdušek O, Šima R, Ayllón N, Jalovecká M, Perner J, de la Fuente J. Interactions of the tick immune system with transmitted pathogens. *Cell Infect Microbiol* 2013;**3**: 1−15.

Downloaded From Vet-eBooks.com

114. Hill CA. *Genome analysis of the major tick and mite vectors of human pathogens*. 2010. Available from: http://www.genome.gov/Pages/Research/DER/PathogensandVectors/Tick_and_Mite_Genomes_Cluster_White_Paper_12Jan2011.pdf.
115. Xu G, Fang QQ, Keirans JE, Durden LA. Molecular phylogenetic analyses indicate that the *Ixodes ricinus* complex is a paraphyletic group. *J Parasitol* 2003;**89**(3):452−7.
116. Oliver J, Bremmer K. Cytogenetics of ticks. III. Chromosome and sex determination in some Australian hard ticks (Ixodidae). *Ann Entomol Soc Am* 1968;**61**:837−44.
117. Goroschenko Y. The karyotypes of argasid ticks of the USSR fauna in connection with their taxonomy. *Tsitologiya* 1962;**4**:137−49.
118. Oliver JH. Cytogenetics of mites and ticks. *Annu Rev Entomol* 1977;**22**:407−29.
119. Oliver JH, Owsley MR, Hutcheson HJ, James AM, Chen C, Irby WS, et al. Conspecificity of the ticks *Ixodes scapularis* and *I. dammini* (Acari: Ixodidae). *J Med Entomol* 1993; **30**(1):54−63.
120. Chen C, Munderloh UG, Kurtti TJ. Cytogenetic characteristics of cell lines from *Ixodes scapularis* (Acari: Ixodidae). *J Med Entomol* 1994;**31**(3):425−34.
121. Munderloh UG, Liu Y, Wang M, Chen C, Kurtti TJ. Establishment, maintenance and description of cell lines from the tick *Ixodes scapularis*. *J Parasitol* 1994;**80**(4):533−43.
122. Hill CA, Guerrero FD, Van Zee JP, Geraci NS, Walling JG, Stuart JJ. The position of repetitive DNA sequence in the southern cattle tick genome permits chromosome identification. *Chromosome Res* 2009;**17**(1):77−89.
123. Meyer JM, Kurtti TJ, Van Zee JP, Hill CA. Genome organization of major tandem repeats in the hard tick, *Ixodes scapularis*. *Chromosome Res* 2010;**18**(3):357−70.
124. Ullmann A., Stuart J., H.C. Tick. Cole C., Hunter W., editors. Heidelberg: Springer-Verlag; 2008.
125. Klompen J, Black 4th W, Keirans J, Norris D. Systematics and biogeography of hard ticks, a total evidence approach. *Cladistics* 2002;**16**:79−102.
126. Ribeiro JM, Alarcon-Chaidez F, Francischetti IM, Mans BJ, Mather TN, Valenzuela JG, et al. An annotated catalog of salivary gland transcripts from *Ixodes scapularis* ticks. *Insect Biochem Mol Biol* 2006;**36**(2):111−29.
127. Klompen JS, Black 4th WC, Keirans JE, Oliver Jr JH. *Evolution of ticks*. 1996. p. 141−61.
128. Dobson SJ, Barker SC. Phylogeny of the hard ticks (Ixodidae) inferred from 18S rRNA indicates that the genus *Aponomma* is paraphyletic. *Mol Phylogenet Evol* 1999;**11**(2):288−95.
129. Klompen J. Comparative morphology of argasid larvae (Acari: Ixodida: Argasidae), with notes on phylogenetic relationships. *Ann Rev Entomol Soc Am* 1992;**85**:541−60.
130. Barker SC, Murrell A. Systematics and evolution of ticks with a list of valid genus and species names. *Parasitology* 2004;**129**(Suppl:S15−36).
131. Mans BJ, de Castro MH, Pienaar R, de Klerk D, Gaven P, Genu S, et al. Ancestral reconstruction of tick lineages. *Ticks Tick Borne Dis* 2016.
132. Navajas M, Fenton B. The application of molecular markers in the study of diversity in acarology: a review. *Exp Appl Acarol* 2000;**24**(10−11):751−74.
133. Meyer J, Hill C. In: Roe M, Sonenshine D, editors. *Tick genetics, genomics and proteomics*. Oxford: Oxford University Press; 2014.
134. Araya-Anchetta A, Busch JD, Scoles GA, Wagner DM. Thirty years of tick population genetics: a comprehensive review. *Infect Genet Evol* 2015;**29**:164−79.
135. Hoogstraal H, Aeschlimann A. Tick-host specificity. *Bull Soc Entomol Suisse* 1982;**55**: 5−32.
136. Black WC, Piesman J. Phylogeny of hard- and soft-tick taxa (Acari: Ixodida) based on mitochondrial 16S rDNA sequences. *Proc Natl Acad Sci USA* 1994;**91**(21):10034−8.

Downloaded From Vet-eBooks.com

137. Crampton A, McKay I, Barker S. Phylogeny of hard ticks (Ixodida) inferred from nuclear ribosomal DNA. *Int J Parasitol* 1996;**26**:511−7.
138. Black WC, Klompen JS, Keirans JE. Phylogenetic relationships among tick subfamilies (Ixodida: Ixodidae: Argasidae) based on the 18S nuclear rDNA gene. *Mol Phylogenet Evol* 1997;**7**(1):129−44.
139. Klompen H, Lekveishvili M, Black WC. Phylogeny of parasitiform mites (Acari) based on rRNA. *Mol Phylogenet Evol* 2007;**43**(3):936−51.
140. Zahler M, Filippova NA, Morel PC, Gothe R, Rinder H. Relationships between species of the *Rhipicephalus sanguineus* group: a molecular approach. *J Parasitol* 1997;**83**(2): 302−6.
141. Fukunaga M, Yabuki M, Hamase A, Oliver Jr JH, Nakao M. Molecular phylogenetic analysis of ixodid ticks based on the ribosomal DNA spacer, internal transcribed spacer 2, sequences. *J Parasitol* 2000;**86**(1):38−43.
142. Tian Z, Liu G, Xie J, Yin H, Luo J, Zhang L, et al. Discrimination between *Haemaphysalis longicornis* and *H. qinghaiensis* based on the partial 16S rDNA and the second internal transcribed spacer (ITS-2). *Exp Appl Acarol* 2011;**54**(2):165−72.
143. Anstead CA, Krakowetz CN, Mann AS, Sim KA, Chilton NB. An assessment of genetic differences among ixodid ticks in a locus within the nuclear large subunit ribosomal RNA gene. *Mol Cell Probes* 2011;**25**(5−6):243−8.
144. Lv J, Wu S, Zhang Y, Zhang T, Feng C, Jia G, et al. Development of a DNA barcoding system for the Ixodida (Acari: Ixodida). *Mitochondrial DNA* 2014;**25**(2):142−9.
145. Delaye C, Aeschlimann A, Renaud F, Rosenthal B, De Meeûs T. Isolation and characterization of microsatellite markers in the *Ixodes ricinus* complex (Acari: Ixodidae). *Mol Ecol* 1998;**7**(3):360−1.
146. Ullmann AJ, Piesman J, Dolan MC, Iv WC. A preliminary linkage map of the hard tick, *Ixodes scapularis*. *Insect Mol Biol* 2003;**12**(2):201−10.
147. Van Zee J, Black WC, Levin M, Goddard J, Smith J, Piesman J. High SNP density in the blacklegged tick, *Ixodes scapularis*, the principal vector of Lyme disease spirochetes. *Ticks Tick Borne Dis* 2013;**4**(1−2):63−71.
148. Van Zee J, Piesman J, Hojgaard A, Black 4th W. Nuclear markers reveal predominately north to south gene flow in *Ixodes scapularis* the tick vector of the Lyme disease spirochete. *PLoS One* 2015;**10**:e0139630.
149. Gulia-Nuss M, Nuss AB, Meyer JM, Sonenshine DE, Roe RM, Waterhouse RM, et al. Genomic insights into the *Ixodes scapularis* tick vector of Lyme disease. *Nat Commun* 2016;**7**:10507.
150. Humphrey PT, Caporale DA, Brisson D. Uncoordinated phylogeography of *Borrelia burgdorferi* and its tick vector, *Ixodes scapularis*. *Evolution* 2010;**64**(9):2653−63.
151. Norris DE, Klompen JS, Keirans JE, Black WC. Population genetics of *Ixodes scapularis* (Acari: Ixodidae) based on mitochondrial 16S and 12S genes. *J Med Entomol* 1996;**33**(1): 78−89.
152. Sakamoto JM, Goddard J, Rasgon JL. Population and demographic structure of *Ixodes scapularis* Say in the eastern United States. *PLoS One* 2014;**9**(7):e101389.
153. Grbić M, Van Leeuwen T, Clark RM, Rombauts S, Rouzé P, Grbić V, et al. The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature* 2011;**479**(7374): 487−92.
154. Cornman R, Schatz M, Johnston J, Chen Y-P, Pettis J. Hunt Gea. Genomic survey of the ectoparasitic mite *Varroa destructor*, a major pest of the honey bee *Apis mellifera*. *BMC Genomics* 2010;**11**:602.

Downloaded From Vet-eBooks.com

155. Chan TF, Ji KM, Yim AK, Liu XY, Zhou JW, Li RQ, et al. The draft genome, transcriptome, and microbiome of *Dermatophagoides farinae* reveal a broad spectrum of dust mite allergens. *J Allergy Clin Immunol* 2015;**135**(2):539−48.
156. Jeyaprakash A, Hoy MA. The mitochondrial genome of the predatory mite *Metaseiulus occidentalis* (Arthropoda: Chelicerata: Acari: Phytoseiidae) is unexpectedly large and contains several novel features. *Gene* 2007;**391**(1−2):264−74.
157. Hoy MA, Waterhouse RM, Wu K, Estep AS, Ioannidis P, Palmer WJ, et al. Genome sequencing of the phytoseiid predatory mite *Metaseiulus occidentalis* reveals completely atomised Hox genes and super-dynamic intron evolution. *Genome Biol Evol* 2016.
158. Rider SD, Morgan MS, Arlian LG. Draft genome of the scabies mite. *Parasit Vectors* 2015;**8**:585.
159. Guerrero F, Nene V, George J, Barker S, Willadsen P. Sequencing a new target genome: the *Boophilus microplus* (Acari: Ixodidae) genome project. *J Med Entomol* 2006;**42**:9−16.
160. Guerrero FD, Moolhuijzen P, Peterson DG, Bidwell S, Caler E, Bellgard M, et al. Reassociation kinetics-based approach for partial genome sequencing of the cattle tick, *Rhipicephalus* (*Boophilus*) *microplus*. *BMC Genomics* 2010;**11**:374.
161. Van Zee J, Schlueter J, Schlueter S, Dixon P, Brito Siera C, Hill C. Gene duplication and genome organization in the *Ixodes scapularis* and other ixodid ticks. *BMC Genomics* 2016.
162. Ayllón N, Villar M, Galindo RC, Kocan KM, Šíma R, López JA, et al. Systems biology of tissue-specific response to *Anaplasma phagocytophilum* reveals differentiated apoptosis in the tick vector *Ixodes scapularis*. *PLoS Genet* 2015;**11**(3):e1005120.
163. Villar M, Ayllón N, Alberdi P, Moreno A, Moreno M, Tobes R, et al. Integrated metabolomics, transcriptomics and proteomics identifies metabolic pathways affected by *Anaplasma phagocytophilum* infection in tick cells. *Mol Cell Proteomics* 2015;**14**(12):3154−72.
164. Grabowski JM, Perera R, Roumani AM, Hedrick VE, Inerowicz HD, Hill CA, et al. Changes in the proteome of langat-infected *Ixodes scapularis* ISE6 cells: metabolic pathways associated with flavivirus infection. *PLoS Neglected Trop Dis* 2016;**10**(2):e0004180.
165. Deleted in review.
166. Schofield CJ, Galvão C. Classification, evolution, and species groups within the Triatominae. *Acta Trop* 2009;**110**(2−3):88−100.
167. Dorn PL, Calderon C, Melgar S, Moguel B, Solorzano E, Dumonteil E, et al. Two distinct *Triatoma dimidiata* (Latreille, 1811) taxa are found in sympatry in Guatemala and Mexico. *PLoS Neglected Trop Dis* 2009;**3**(3):e393.
168. de la Rua N, Stevens L, Dorn PL. High genetic diversity in a single population of *Triatoma sanguisuga* (LeConte, 1855) inferred from two mitochondrial markers: cytochrome b and 16S ribosomal DNA. *Infect Genet Evol* 2011;**11**(3):671−7.
169. Bargues MD, Klisiowicz DR, Gonzalez-Candelas F, Ramsey JM, Monroy C, Ponce C, et al. Phylogeography and genetic variation of *Triatoma dimidiata*, the main Chagas disease vector in Central America, and its position within the genus *Triatoma*. *PLoS Neglected Trop Dis* 2008;**2**(5):e233.
170. Giordano R, Cortez JC, Paulk S, Stevens L. Genetic diversity of *Triatoma infestans* (Hemiptera: Reduviidae) in Chuquisaca, Bolivia based on the mitochondrial cytochrome b gene. *Mem Inst Oswaldo Cruz* 2005;**100**(7):753−60.
171. Monteiro FA, Perez R, Panzera F, Dujardin JP, Galvao C, Rocha D, et al. Mitochondrial DNA variation of *Triatoma infestans* populations and its implication on the specific status of *T. melanosoma*. *Mem Inst Oswaldo Cruz* 1999;**94**(Suppl. 1):229−38.

Downloaded From Vet-eBooks.com

172. Almeida CE, Pacheco RS, Haag K, Dupas S, Dotson EM, Costa J. Inferring from the Cyt B gene the *Triatoma brasiliensis* Neiva, 1911 (Hemiptera: Reduviidae: Triatominae) genetic structure and domiciliary infestation in the state of Paraiba, Brazil. *Am J Trop Med Hyg* 2008;**78**(5):791−802.
173. Monteiro FA, Donnelly MJ, Beard CB, Costa J. Nested clade and phylogeographic analyses of the Chagas disease vector *Triatoma brasiliensis* in Northeast Brazil. *Mol Phylogenet Evol* 2004;**32**(1):46−56.
174. Bargues MD, Klisiowicz DR, Panzera F, Noireau F, Marcilla A, Perez R, et al. Origin and phylogeography of the Chagas disease main vector *Triatoma infestans* based on nuclear rDNA sequences and genome size. *Infect Genet Evol* 2006;**6**(1):46−62.

Downloaded From Vet-eBooks.com

Multilocus Sequence Typing of Pathogens: Methods, Analyses, and Applications

16

M. Pérez-Losada [1,2,3], *M. Arenas* [4,5], *E. Castro-Nallar* [6]
[1]George Washington University, Ashburn, VA, United States; [2]Universidade do Porto, Vairão, Portugal; [3]Children's National Medical Center, Washington, DC, United States; [4]University of Porto, Porto, Portugal; [5]Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP), Porto, Portugal; [6]Universidad Andrés Bello, Santiago, Chile

1. Introduction

The ability to accurately distinguish between strains of infectious pathogens is crucial for efficient epidemiological and surveillance analysis, studying microbial population structure and dynamics and, ultimately, developing improved public health control strategies.[1] To further such general goals, several molecular typing methods have been proposed that can identify isolates worldwide (global epidemiology) and/or in localized disease outbreaks (local epidemiology); see Foley for a review.[2] Nonetheless, since 1998, the established standard for molecular typing is multilocus sequence typing[3] (MLST). MLST was built on the well-established population genetic concepts and methods of the multilocus enzyme electrophoresis (MLEE) technique, but provides significant advantages over this and other typing approaches (see Section 4 for advantages and caveats). MLST examines nucleotide variation in sequences of internal fragments of usually seven housekeeping genes: that is, those encoding fundamental metabolic functions (see Section 2 for molecular design and development of MLST). For each gene, the different sequences present within a species are assigned as distinct alleles and, for each isolate, the alleles at each of the seven loci define the allelic profile or sequence type (ST). Each isolate is therefore unambiguously characterized by a series of seven integers, which correspond to the alleles at the seven housekeeping loci. Most bacterial species have sufficient variation within housekeeping genes to provide many alleles per locus, allowing billions of distinct allelic profiles to be distinguished using just seven loci. Alternatively, isolate identification and tracking can be performed using the nucleotide data directly, although this approach is more frequently used for population studies (see Section 5 for methods of analyses).

MLST is widely used for molecular typing.[4–7] Numerous examples exist of their use for describing the population structure of pathogens, vaccine studies, tracking transmission of epidemic strains, and identifying species and virulent strains associated with disease (see Section 6 for applications). This was made possible by three

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00016-0
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

improvements in molecular microbiology[4] involving: (1) bacterial evolution and population biology knowledge (discussed later); (2) high-throughput nucleotide sequencing (see Section 2 for molecular basis); and (3) internet databases (see Section 3). The bacterial population studies undertaken from the 1980s onward were central to the development of MLST. Those studies showed that genetic exchange among bacteria was more common than previously thought, leading to a reassessment of the role of sexual processes in the structuring of bacterial populations. Using sequence data, it has been shown that recombination (mosaic genes) was frequent not only in genes under diversifying selection (e.g., antigen-encoding and antibiotic-resistant determinant genes), but also in genes under purifying selection (housekeeping genes). This suggested that the clonal model (variation can only arise by mutation) was not universal and led to the proposal of new nonclonal or panmictic (variation is mainly generated by recombination) and partially clonal models of bacterial population structure. Consequently, typing methods needed to accommodate a broader spectrum of population structures and be able to distinguish among them, hence providing not only discriminatory power but also information about the clonal structure of the organism under study. Therefore, only molecular techniques that can contrast results across independent markers (such as MLST) would be adequate for bacterial typing and population genetic inferences.

In the following sections, we describe in more detail all the epigraphs mentioned in this introduction. We also refer the reader to other reviews on MLST for complementary information.[1,4–6,8–13]

2. Molecular Design and Development of Multilocus Sequence Typing

The principal element in the design of an MLST scheme is the choice of genetic loci. The selection and number of loci is based on principle, precedent, and practice. Since MLST was developed as an updated version of MLEE, which indexes variation of multiple core metabolic or housekeeping genes at the protein level, the selected loci typically correspond to housekeeping genes encoding proteins for core metabolic functions. Furthermore, housekeeping genes are expected to be somewhat conserved and vertically transmitted and thus should reveal genetic relationships among strains without concern for the influence of host or environmental factors. For instance, such influences might occur when genes encoding hypervariable surface proteins are subject to immune-driven diversifying selection or genes under antibiotic selection. The genes should be physically spaced around the genome in order to minimize genetic linkage of loci.

As a matter of principle and practicality, multiple loci of sufficient length need to be surveyed in order to provide a high level of discrimination. The first MLST scheme was designed by Maiden and colleagues[3] and included six, later expanded to seven, loci. Most investigators have followed this precedent and developed schemes of seven loci. The length of nucleotide sequence amplified for each locus is generally in the

Downloaded From Vet-eBooks.com

range of 400−600 bp and is determined largely by the parameters of automated sequencing instruments available at the time the first MLST scheme was developed in 1998. Most MLST nucleotide sequence data are generated by the Sanger sequencing method, however, high-throughput technologies such as pyrosequencing,[10,14] sequencing-by-synthesis, and single-molecule sequencing[5,15] will likely be the methods of choice in the future for both targeted-amplicon and whole-genome sequencing. Those technologies are capable of generating accurate read lengths of ~500 bp to 10 kb (PacBio RS II and Sequel systems) and up to 25−50 million paired-end reads (Illumina MiniSeq/MiSeq platforms) per run. Moreover, the design of barcoded primers allows simultaneous and efficient sequencing of homologous products from hundreds of samples in the same run[16]; see also www.pacb.com and Chen et al.[15]

The development of a new MLST scheme from scratch involves four initial steps (Table 16.1): (1) identification of loci, (2) PCR primer design, (3) survey of a small number of representative strains, and (4) analysis of nucleotide sequence data to

Table 16.1 Stages in the Design of an MLST Scheme

Actions	Criteria
Analyze reference genome to identify 12−18 candidate loci	• Single-copy gene • Putative core housekeeping gene Genes evenly spaced in genome
Design nested PCRs using primer select software	• Outer PCR product about 1000−1500 bp • Inner PCR product about 400−600 bp
Select 20−25 representative strains	• Isolated in different years and different geographic sites • No known epidemiological linkage by transmission or shared phenotypic characteristics
Perform nested PCRs of the 20−25 strains and redesign primers as needed	
Analyze nucleotide sequence data	• Rank loci by level of nucleotide polymorphisms and select 7−9 loci with high level of polymorphism
Select 75−100 strains using the previously mentioned criteria, type using the 7−9 loci, and perform analysis	• Confirm loci are under purifying selection • Assign each unique sequence an allele number • Assign each isolate an ST • The greater the number of STs, the greater the discriminatory power of the MLST

Downloaded From Vet-eBooks.com

establish neutral evolution of loci and level of strain discrimination. For many bacterial species, the selection of loci is greatly aided by the availability of annotated whole genomes, which allows ready identification of housekeeping genes and their physical location in the genome. An absolute requirement for loci included in an MLST scheme is that there is only a single copy of the gene in the genome. It is advisable to choose more than seven loci because not all loci will pass subsequent tests of utility, and typically 12−18 loci are selected for subsequent tests. As much as possible, the loci should be evenly spaced across the genome and certainly separated by several tens of thousands of base pairs, although no rules allow a precise estimate of the maximum size of bacterial genomic fragments that can undergo recombination. The physical location of loci within genomes may differ among strains, so use of a single reference strain, which is often all that is available, is at best an approximation. The design of primers is greatly assisted by the availability of open access and commercial software for primer design but ultimately depends on trial and error.[17,18] Most MLST schemes use a nested PCR design both to increase sensitivity for samples with a low bacterial DNA copy number and, more importantly, to provide a high-quantity and high-quality PCR product for sequencing. The initial evaluation of candidate loci is most easily accomplished with a small number of strains (20−25), and the strains should not be epidemiologically linked or share defining characteristics, such as antibiotic resistance, that might lead to over-sampling of a clonal population. Temporally and geographically separated strains provide one likely basis for accomplishing this goal. The data from this small set of strains should allow the stratification of loci on the basis of efficiency of detection by nested PCR and level of genetic variation. They also provide the opportunity to optimize primer design.

At least 7−9 loci that could be amplified from all test strains and showed a reasonably high level of genetic diversity[5,19] should then be evaluated with a larger data set of 70−100 strains to accomplish the initial data analysis for evolutionary neutrality and level of strain discrimination. The same rules for selection of strains apply here as mentioned earlier. A representative collection of strains should be used, but in practice it is only possible to avoid obvious pitfalls, such as selecting strains from a known outbreak. The purpose of the initial analysis of MLST data is to confirm that the chosen loci are under purifying selection, to assess the level of polymorphism at each locus, and to determine whether a sufficient level of discrimination is achieved for epidemiological studies. The number of unique nucleotide sequences among the 70−100 strains tested establishes the level of polymorphism, and alleles that are the most polymorphic will provide the greatest degree of discrimination among strains. While low levels of polymorphism are a reason to reject an allele for inclusion in an MLST scheme because they will provide little discriminatory power, the seven most polymorphic alleles are not necessarily the best choice. Ideally, all seven loci will contribute equally to the discriminatory power of the method and a very high level of variation may be indicative of diversifying selection pressure. On the other hand, evolutionary neutrality is a desirable, but not absolutely necessary, characteristic of loci used in an MLST typing scheme. In fact, most other methods for strain typing use highly polymorphic loci, which are often known to be subject to selection pressure. If one or more of the initially selected loci fail the test of neutrality, or no

Downloaded From Vet-eBooks.com

combination of 6−7 loci provides sufficient strain discrimination, other loci surveyed in the test set can be evaluated with the larger data set, and a new 6−7 loci MLST scheme can be designed. Finding the right balance in terms of efficiency of PCR amplification, locus neutrality, strain discrimination, and comparability of polymorphisms across loci is ultimately a matter of judgment rather than the application of precise rules.

Once candidate loci have been chosen and the MLST scheme defined, application of the method in the context of epidemiological studies will establish its reliability in typing large numbers of diverse strains and its ability to provide sufficient strain discrimination to address epidemiological questions of interest. For strains that cannot be typed using the initial PCR primers, it is generally easy to design new primers. Although the choice of loci used in the MLST scheme could be modified as more strains are typed (e.g., to increase discrimination), one of the strengths of MLST as a typing method would be sacrificed; namely, the comparability of data generated over time and by multiple investigators. Because the sequence type or ST is defined by the set of distinctly numbered alleles at the seven loci, changing loci would result in new STs that could not be directly compared to STs defined using the previous MLST scheme. In that regard, using in silico MLST approaches based on whole-genome data allows us to compare different typing schemes for the same group or even integrate genomic inferences with information-rich MLST databases.[20−23]

If an epidemiological study requires discrimination of closely related strains, as may be necessary to examine short-term transmission of antibiotic-resistant isolates, rather than add to or change the loci in an MLST scheme, a better strategy is to supplement MLST with additional highly polymorphic markers, such as genes encoding antigens, cell surface proteins, ribosomal genes, or tandem repeats.[11,19,24−26]

Over the last few years, other typing approaches have been developed based on similar principles as MLST. Multilocus Variable number of tandem repeats Analysis (MLVA) uses polymorphic repeated sequences (VNTR) instead of housekeeping genes. Comparative studies between MLVA and MLST have yielded similar results, for example, van Cuyck et al.,[27] and in recently originated species, the MLVA approach may have higher discriminatory power.[28] Similarly, the Ribosomal Multilocus Sequence Typing method (rMLST) has also been proposed to index the molecular variation of 53 genes encoding bacterial ribosome protein subunits.[11] This method pursues the integration of a taxonomic and typing method in a similar curated MLST scheme. Although more expensive, the rMLST is likely to provide better resolution than previous methodologies. Likewise, core-genome (cg) MLST has been developed to overcome lack of resolution of MLST schemes of certain taxa. By collecting a sample of genome sequences representing extant diversity, the cgMLST scheme uses >1000 genes to create sequence types that provide increased resolution for clonal populations of bacteria.[29] Finally, in order to achieve even greater resolution, other approaches have been developed based on core/accessory genes or distributed genes among bacterial species that have the same MLST profile.[30,31] This new approach could skip the laborious and time-consuming steps needed to develop bacteria-specific MLST schemes.

Downloaded From Vet-eBooks.com

3. Multilocus Sequence Typing Databases

One of the goals of the MLST approach was the development of online platforms containing MLST databases to which public health officials and researchers could both have access and contribute; and from which clinical, epidemiological and population studies could benefit.[3,4,8] The first MLST websites were based on single databases implemented in the MLSTdB software[32]; but as MLST schemes began to expand, several limitations became apparent: redundant information (each record contained the ST designation and the allelic profile), isolate bias (single databases were dominated by specific studies), and access (all databases were stored at a single location). To overcome these limitations, a new network-based database software, MLSTdBNet,[33] was developed and implemented on the PubMLST site (http://pubmlst.org/). This site is served by two databases: (1) a profiles database that contains the sequences of each MLST allele for each locus linked to an allele number, and (2) an allelic profiles database with their ST designations. The profile database can then serve other isolate databases. For each scheme on the PubMLST site there is a PubMLST isolate database that aims to include at least one isolate for each ST. MLST databases are hence different from other depository databases, such as GenBank, not only in organization but also in that they are actively curated for accuracy. It is important to highlight that MLST databases do not embody the global diversity of an organism but the extent of its diversity at the time they are accessed. Moreover, stored data is unstructured and does not necessarily represent natural populations either. As high-throughput sequencing becomes more affordable, PubMLST is increasingly including whole-genome sequences, for example, BIGSdb.[34]

Several other websites are accessible through the PubMLST site. The PubMed (NCBI) is linked to PubMLST databases, so original publications describing MLST schemes can be retrieved. The AgdbNet—antigen sequence database software for bacterial typing[35]—is also integrated into the system. Other websites are available for the storage and access of MLST data. At the time of writing, 93 MLST schemes (82 for bacteria, 9 for eukaryotes, and 1 each for plasmids and bacteriophage) could be accessed via the PubMLST site. The PubMLST primary site is also mirrored in four locations, three in UK and one in Pittsburg (USA). This provides access to MLST data globally and assures that databases are stored in multiple locations. A detailed description of the MLST databases, their structure, and most of the published MLST schemes can be found in Maiden.[4]

Other websites (www.spatialepidemiology.net/ and beta.mlst.net/Instructions/mlstmaps.html) have also been developed that incorporate geospatial information in bacterial epidemiological studies. Those websites provide precise locality data related to strain distribution and a map-based interface for displaying and analyzing epidemiological information. Moreover, the portal www.eMLSA.net enables species identification by means of a taxonomic platform. The integration of genomic and epidemiological data together with geographic information through MLST databases will greatly improve our ability to track and prevent infectious pathogens and associated diseases.

Downloaded From Vet-eBooks.com

4. Advantages and Disadvantages of Multilocus Sequence Typing

As the number of schemes available has increased, MLST has become the most commonly used method of pathogen typing. In comparison to older methods (serotyping; MLEE), the use of genetic variation gives MLST the advantage of producing variable data (more resolution) that are universally comparable (within schemes), easily validated, and readily shared across laboratories. The use of generic sequencing technology makes MLST a broadly applicable methodology that can be fully automated and scalable from single isolates to thousands of samples. Because the materials needed for MLST analysis—DNA or dead cells—are easily transported among laboratories without the problems associated with infective materials, both the biological samples and the resulting data are highly portable. Furthermore, the use of online electronic databases (see Section 3) to store and curate MLST schemes makes them a globally accessible resource.

MLST targets variation at multiple housekeeping loci. The number of loci that need to be evaluated to confidently assign an ST has been minimized to reduce the expense and time required for characterization, with most studies using 6−10 loci. If performed manually, evaluating even this many loci can be time consuming. However, fully automated systems, for example, robotics[36] provide a high-throughput pipeline for data collection that can run large volumes of samples with increased reliability. Likewise, commercial solutions, such as Ion Torrent AmpliSeq panels targeting MLST schemes, can reduce costs down to cents per marker (www.ampliseq.com). As sequencing technology progresses, we expect the cost of automation to decrease, so data interpretation, rather than data generation, will be the likely limiting factor in our understanding of pathogen population dynamics.

By focusing on sequence variation, MLST provides a highly replicable and reproducible typing method. Additionally, the focus on housekeeping genes provides significant amounts of genetic data that can be used to calculate pathogen population genetic parameters (see Section 5) at both local and global scales. Those parameters can be then used to construct more sophisticated models of pathogen evolution and epidemiology that will improve our understanding of how to control the spread of disease. However, there is no single set of universal housekeeping genes that can be used for all pathogens as the recombination rates, substitution rates, and levels of selection vary across loci and species.[13] Therefore, a unique set of loci must be identified for each novel, untyped pathogen under study. The rapid increase of available microbial genomes will make data mining for housekeeping genes more feasible, reducing the time and cost required for constructing new MLST schemes.

Currently, the main drawback of the MLST method is that the selection of housekeeping loci requires reference genomes.[37] Moreover, not all pathogens are suitable for MLST methods. Some pathogens (e.g., *Mycobacterium tuberculosis* and *Yersinia pestis*) exhibit very little variation throughout their entire genome, most likely representing "evolutionarily young" pathogens that have not yet accumulated sufficient genetic variation to differentiate strains. For typing these pathogens, more rapidly

Downloaded From Vet-eBooks.com

evolving loci (e.g., insertion sequences or antibiotic-resistance determinants) or more markers (genome-wide SNPs) are needed. Conversely, some bacterial genomes have accumulated so much variation that MLST housekeeping genes do not provide adequate information for typing. As we advance MLST schemes in the postgenomic era, we should be able to combine information-rich and widely adopted schemes with cost-effective whole-genome sequencing.

5. Analytical Approaches

There are two basic strategies to the analysis of MLST data (Fig. 16.1), one relies on allele and ST designations to estimate relatedness among isolates (*allele-based methods*), and so ignores the number of nucleotide differences between alleles; and the other relies on nucleotide sequences directly to estimate relatedness and population parameters (*nucleotide-based methods*). The allele-based approach has been adopted from the analysis of MLEE data and so methods based on this strategy were the first applied to the analysis of MLST data.[3,38] The allele-based approach is thought to work well in nonclonal organisms (e.g., *Helicobacter pylori*), while nucleotide-based

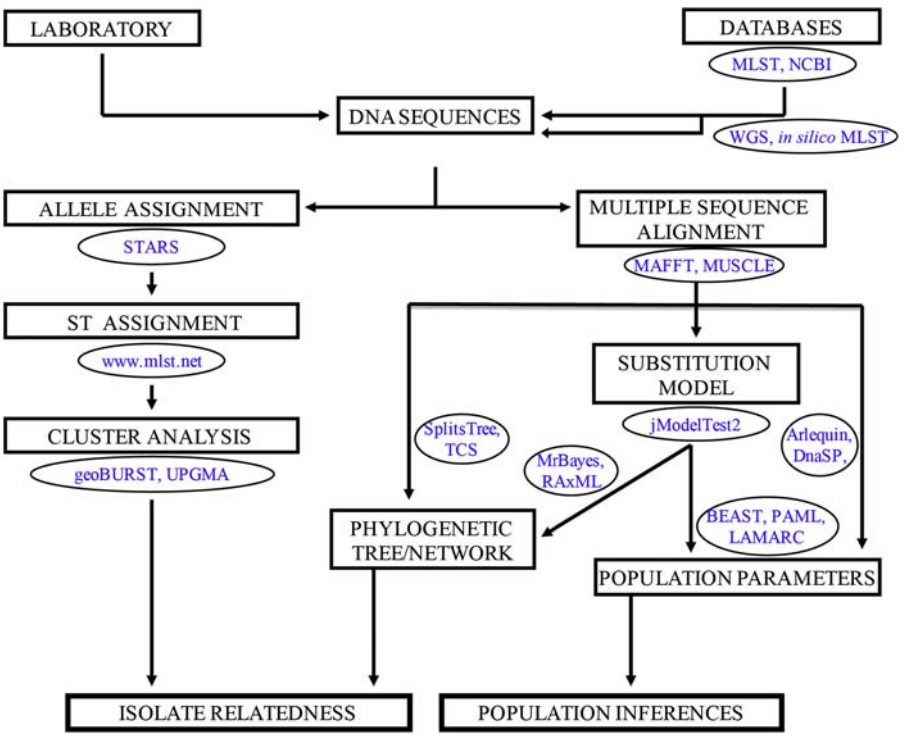


Figure 16.1 Pipeline showing data and tasks (*boxes*) and databases and computer programs (*circles*) commonly used in the analysis of MLST data.

Downloaded From Vet-eBooks.com

approaches are preferable for clonal organisms (e.g., *Escherichia coli*) since the former are likely misleading.[4] But in practice, most microbes show some degree of clonality (clonal complex) in their populations; hence, in our opinion, both types of analyses should be conducted in population and epidemiological studies, for example, Loubna et al.[39] In this section, we present a brief description of some of the most commonly used approaches for analyzing MLST data. We refer the reader to previous reviews for a more detailed description.[9]

5.1 Allele-Based Methods

Since alleles are the unit of analysis, all these methods first require assigning an allele number to each DNA sequence from each locus. This is done by matching our sequences against those stored in public MLST databases (see Section 3). If no match is found, a new number is assigned in order of discovery. Several computational programs have been developed for this task, although Sequence Typing Analysis and Retrieval System (STARS) seems to be very functional and widely popular.[9] The STARS interface was specifically designed for typing and allows the assembly of large number of sequences at once.

Once alleles have been assigned, data are entered in the MLST websites to acquire an ST profile. At this point, exploratory analysis (e.g., allele and profile frequencies, polymorphism estimates, and codon usage) of the data can be performed. The software package Sequence Type Analysis and Recombinational Tests (START2) can perform all these tasks.[40] Relatedness among STs can be then displayed using methods of cluster reconstruction, such as the Based Upon Related Sequences Types (eBURST) approach and the simple Unweighted Pair Group Method with Arithmetic Mean (UPGMA). eBURST[41] is based on a simple model of clonal expansion and diversification. It first identifies mutually exclusive groups of related STs and attempts to identify the founding ST of each group. Bootstrap estimates are also calculated to assess confidence in the groupings. The algorithm then predicts the descent from the predicted founding ST to the other STs in the group, displaying the output as a radial diagram, centered on the predicted founding ST. A globally optimized version (goeBURST) is also available that identifies alternative patterns of descent using a graphic matroid approach.[42] In 2012, a new approach (PHYLOViZ) was released for microbial epidemiological and population analysis that allows for the integration of allelic profiles from MLST or MLVA methods (although Single Nucleotide Polymorphism data can also be included) and associated epidemiological data.[43] PHYLOViZ uses goeBURST for representing the possible evolutionary relationships between strains.

The traditional UPGMA method relies on a matrix of distances to estimate isolate relatedness. Distances are calculated for each pair of STs based on the number of allele differences, and groups are then sequentially clustered in order of similarity (i.e., allelic matches). Additional distance and parsimony methods have been proposed to estimate relatedness based on allele frequencies, but note that distance methods generally outperform parsimony methods.[44]

Downloaded From Vet-eBooks.com

Allele-based methods have the advantage of simplicity and speed, which are crucial for efficient epidemiological surveillance and public health management, but disregard much of the evolutionary information contained at the nucleotide level. A larger and more sophisticated plethora of nucleotide-based methods exist to estimate isolate relationships and key population parameters.

5.2 Nucleotide-Based Methods

Any analysis of nucleotide data usually begins with a multiple sequence alignment (MSA) (i.e., estimation of the homologous nucleotide sites). Since the loci used for MLST usually evolve very slowly and code for proteins, this step becomes trivial, particularly at the amino acid level. If needed, several fast and accurate iterative aligning strategies are implemented in MAFFT[45] and MUSCLE.[46]

Once an alignment has been generated, we have to determine the model of evolution that fits the data the best. Model choice is a critical issue and the implemented model (or lack thereof) will affect all subsequent phylogenetic[47] and population analyses (following two sections). This issue is usually assessed within a phylogenetic framework, see Posada et al.[48] Since mid-1990s substitution models have increased in complexity, as parameters reflecting new information on nucleotide substitution processes are added to candidate models.[49] Furthermore, model selection can consider confidence sets of models (model averaging).[48] Several criteria have been proposed for choosing models, such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Decision Theory (DT), and Hierarchical Likelihood Ratio Test (hLRT).[50] Although AIC is the most broadly used method for evaluating model fit, BIC and DT should be preferred.[51] These strategies are implemented in the well-established program jModelTest2.[50]

5.2.1 Phylogenetic Relatedness

Phylogenetic reconstruction methods can be divided into two types, those that proceed algorithmically through distances, for example, UPGMA and neighbor joining (NJ) and those based on optimality criteria. Here, we focus on those that implement maximum likelihood and Bayesian optimality criteria and allow for the implementation of multiple data partitions each under its best-fit model. We find this feature particularly important for analyzing MLST data.

Maximum likelihood (ML) inference attempts to identify the topology that explains the evolution of a set of aligned sequences under a given substitution model of evolution with the greatest likelihood.[52] RAxML[53] implements the ML criterion efficiently and accurately and can handle large data sets of >1000 sequences with >20 kb.[54] Confidence in the estimated relationships (i.e., clade support) is usually assessed using a nonparametric bootstrap procedure,[55] which must be repeated >1000 times to achieve reasonable precision. RAxML can also rapidly estimate bootstrap proportions. Another well-established ML framework is PhyML,[56] which can internally optimize diverse evolutionary parameters.

Although similar to ML inference, Bayesian inference (BI) combines the prior probability of a phylogeny with the likelihood to produce a posterior probability

Downloaded From Vet-eBooks.com

distribution of trees, which can be interpreted as the probability of those trees (or tree) being correct.[57] Clade support is estimated by summarizing this distribution of trees through consensus analysis. Bayesian phylogenies are estimated using Metropolis-coupled Markov chain Monte Carlo (MCMC) methods and both are implemented in programs such as MrBayes.[58] The output of the BI analysis must be evaluated to assure the MCMC chains have mixed well and converged; such tasks can be performed in Tracer.[59] Importantly, the best fitting model can vary across sites. For this reason, programs such as RAxML or MrBayes implement partition-specific (i.e., sites or genomic regions) models that can improve the accuracy of phylogenetic inferences.[60]

Often, gene trees differ even when sampled from the same population. This can be the result of molecular processes (e.g., recombination) or stochastic variation (e.g., lineage sorting). Whatever the case, one may want to check if individual gene topologies are significantly different since ignoring these processes may lead to biased parameter inferences.[61] Multiple ML topological tests have been developed for such purpose and several are implemented in CONSEL.[62]

New coalescent approaches have been developed to deal with stochastic variation in gene trees from multilocus molecular data and to estimate gene trees and species tree. Among such, BEST[63] and *BEAST[64] consider the effect of incomplete lineage sorting (ILS) by implementing the multispecies coalescent model into a Bayesian hierarchical model. When estimating evolutionary relationships among microbes using DNA sequences, the reticulating impact of recombination becomes a significant issue. If recombination is substantial, the evolutionary history of those sequences no longer fits a bifurcating model as those described before, and therefore a tree representation may fail to accurately portray a reasonable genealogy.[65] Under such circumstances, network approaches[66] can be used instead. Recently, Woolley et al.[67] have revised the most common algorithms for building phylogenetic networks and concluded that the union of maximum parsimonious (UMP) trees[68] performed the best. TCS[69] and SplitsTree[70] also performed well at estimating network gene genealogies. Finally, Didelot and Falush[71] have developed a Bayesian coalescent approach (ClonalFrame) that also takes homologous recombination into account while inferring clonal relationships between the members of a sample.

5.2.2 Population Dynamics

The evolution of DNA sequences in natural populations can be described with parameters such as recombination, mutation, growth, and selection rates. Indeed, the accurate estimation of these parameters is key for understanding the dynamics and evolutionary history of those populations, their epidemiology, the potential for and mode of evolution of antibiotic resistance, and ultimately for applying efficient public health control strategies. Population parameters are more efficiently estimated using explicit statistical models of evolution, such as the coalescent approach, hence here we describe some population parameter estimators based on such models.

Recombination is generally defined as the exchange of genetic information between two nucleotide sequences. It influences biological evolution at many different levels as well as affects the estimation of other parameters. Comprehensive assessment of statistical methods for detecting and estimating recombination rates were presented in

Downloaded From Vet-eBooks.com

Martin et al.[72] and Posada et al.[73] These studies concluded that one should not rely on a single method to detect or estimate recombination. With this idea in mind, software packages such as RDP4[74] have been developed to implement a variety of methods for the same data set. RDP4 is a package that includes 12 recombination estimators and allows the user to draw conclusions based on the outcome of multiple tests. Another ML method to detect recombination is GARD,[75] which outperformed previously developed methods. In addition, programs such as LAMARC, LDhat, CodABC, and OmegaMap[76] (described in Pérez-Losada et al.[77]) can be used to estimate recombination rates and, therefore, to quantify the amount of observed recombination. Similarly, these methods can estimate genetic diversity, the most important population parameter. Reviews of classical and newer statistical methods for estimating genetic diversity have been published elsewhere.[78–81]

Another key parameter for characterizing microbial population dynamics is the growth rate, which reflects the variation of genetic diversity over time. Growth rates can be estimated under a certain demographic model (e.g., exponential) or without dependence on a prespecified model, for example, Skyride.[82] The latter approach is implemented in BEAST,[83] which also allows for the analysis of temporally spaced sequence data. Exponential growth rates and genetic diversity can also be estimated in LAMARC.

The standard method for estimating selection in protein-coding DNA sequences is through the nonsynonymous (d_N) to synonymous (d_S) amino acid substitution ratio d_N/d_S (ω). $\omega > 1$ indicates adaptive or diversifying selection, $\omega < 1$ purifying selection, and $\omega \approx 1$ lack of selection (neutral evolution). ω is usually estimated within an ML phylogenetic framework and assuming an explicit model of codon substitution. Such models can be very complex, allowing, for example, ω to vary across amino acid sites and/or tree branches, for example, Yang.[84] If significant evidence (usually obtained through likelihood ratio tests, LRT) of adaptive selection is obtained, then Bayesian tests can be applied to detect amino acid sites under selection, for example, Yang et al.[85] Such methods are implemented and described in more detail in the software package PAML.[84] However, if recombination is suspected in the data, it should be considered when estimating ω to avoid false positively selected sites.[61] Thus, one could estimate recombination and selection rates simultaneously with OmegaMap or CodABC, or account for the former while estimating the latter, for example, HYPHY.[86]

Other key factors in microbial dynamics are time of emergence (e.g., pathogen outbreaks) and geographic distribution of pathogens. New probabilistic models were developed within the Bayesian framework[87] for inference and hypothesis testing of divergence times, ancestral locations and historical patterns of migration (i.e., phylogeographical history). Such models are implemented in BEAST and SPREAD[88] and visualized using virtual globe software, such as Google Earth; they have already begun to be applied to the analysis of MLST and/or genome and SNP data.[89,90]

Most of the nucleotide-based methods described earlier, and others, have been implemented in user-friendly web servers, such as CBSU (cbsuapps.tc.cornell.edu), CIPRES (www.phylo.org), Datamonkey (www.datamonkey.org), or PhyML (www.atgc-montpellier.fr/phyml/).

Downloaded From Vet-eBooks.com

6. Applications of Multilocus Sequence Typing

MLST analysis and databases are standardized and broadly used, filled with historical information, and firmly established in molecular and clinical laboratories worldwide. Consequently, new typing applications are seeking to integrate existing MLST schemes with whole-genome shotgun data to characterize microbial populations, rather than creating from scratch new typing methods. MLST is probably the most flexible typing method since it can be implemented in small laboratories with standard equipment (PCR + Sanger sequencing), as well as in medium-sized facilities with vanguard infrastructure (targeted-amplicon sequencing; AmpliSeq panels, robotics, and so on) or laboratories with whole-genome sequencing capability (in silico MLST).

Although primarily developed for the characterization of organisms (typing), MLST sequence data have also been applied to other endeavors such as molecular epidemiology (e.g., disease transmission and surveillance programs) and public health (e.g., monitor and evaluate vaccination programs), as well as to other areas such as phylogenetics, taxonomy, speciation, population genetics, biosafety, and even to the inference of human migrations.

6.1 Molecular Epidemiology and Public Health

MLST has gained widespread popularity as a typing method and its use has advanced understanding of bacterial evolution and has provided insights into the epidemiology of bacterial diseases. In the context of surveillance and management of disease outbreaks, being able to quickly type and track infectious diseases is of paramount importance. Many studies exemplify the use of MLST in these circumstances: emergence of zoonosis,[89,90] detection of disease outbreaks,[91,92] estimation of prevalence rates,[93,94] and the origins of virulence factors (vertical or horizontally transmitted).[95,96]

MLST data have been also used to infer population structure and study the emergence and spread of antibiotic resistance.[97] For example, MLST has been used to diagnose human-associated population structure in the opportunistic pathogen *Ochrobactrum anthropi*. Romano et al.[98] developed an MLST scheme for this pathogen and used the evolutionary information inherent in the DNA sequences to identify a human-associated subpopulation from their collection of clinical and environmental isolates. Likewise, MLST has been used to track drug-resistance variants through patients. Oteo et al.[99] collected 162 isolates of *Klebsiella pneumoniae* from five hospitals in Spain and used the MLST data to demonstrate the spreading of this bacteria as pathogen and colonizer of newborns and adult patients with multilocus resistance acquired through recombination. Similarly, Lee et al.[100] used MLST to identify epidemic and virulent ciprofloxacin-resistant *E. coli* clones and their population structure in Korea causing urinary tract infections.

In a number of studies, MLST data have been used to reveal the epidemiological history of infectious diseases. For example, MLST has been successful in identifying clinically important strains of *Neisseria meningitides*, that is, hyperinvasive lineages.[101] MLST has been applied to a number of clinically important bacterial populations, including hospital-acquired strains of *Enterococcus faecalis* and

Downloaded From Vet-eBooks.com

Enterococcus faecium,[102,103] and *Streptococcus pneumoniae* strains associated with invasive disease.[104] In some cases, MLST has failed to distinguish clinically relevant populations. For example, *Staphylococcus aureus* isolates from persons with nasal carriage, community-acquired pneumonia, and hospital-acquired invasive disease are evenly distributed among clonal complexes.[105] Similarly, there is a poor correlation between MLST data and tissue tropisms (throat or skin) of *Streptococcus pyogenes* isolates.[106] For phenotypes that are based on one or a few genes, such as antibiotic resistance, correlations with MLST data have been large. The evolutionary history of methicillin-resistant *S. aureus* (MRSA) has been clarified by MLST data, including the typing of the methicillin-resistance genetic element, SCCmec.[107] Along the same lines, MLST has been used to identify transmission chains as demonstrated by Choudhury et al.[108] where the authors identified outbreak sources and characterized outbreaks of gonorrhea. They typed consecutive *gonococcal* strains from London STI clinics over a 9-month period. Clusters of patients with the same strain showed similarities in behavioral and demographic features, suggesting that different strain clusters represent localized transmission chains.

New phylogenetic coalescent models have been developed allowing researchers to infer from genetic data more familiar parameters, such as the reproductive number of viruses,[109] as well as to model epidemiological dynamics that describe changes in population size or date of origin.[110–113] Lastly, examples of MLST and whole-genome sequencing integration abound (see Pérez-Losada et al.[5]). In molecular epidemiology, studies since 2010 combine MLST data with in silico MLST in an effort to put new isolates in context without losing the resolution and insight gained by having the full genetic complement of the bacteria in question.[114–117]

6.2 Species Diagnosis and Phylogenetics

MLST data have been used to distinguish similar species, to inform the division of a genus into species, and to ask whether bacterial species exist. The MLST data are especially useful for species diagnoses as they provide both genealogical information as well as information on recombination.[118] Indeed, even when the MLST are not as discriminating as other approaches, the phylogenetic information available through MLST provides novel insights into species and strain relatedness that impact public health decisions. In a study of *Clostridium difficile*, for example, Marsh et al.[28] found MLST less discriminatory compared to MLVA or restriction endonuclease analysis (REA) although concordant, but the combination of MLST with MLVA provided novel insights into the origins and evolutionary relationships bearing clinical and public health importance. Similarly, a phylogenetic analysis of concatenated sequences of seven MLST loci for *Bacillus pseudomallei* and *Bacillus thailandensis*, both soil saprophytes, and *Bacillus mallei*, the cause of glanders, showed that all *B. pseudomallei* strains were tightly clustered and well resolved from all *B. thailandensis* strains.[119] However, *B. mallei* clustered with *B. pseudomallei* and, although designated as a "species," can be considered to be a strain (or clone) of *B. pseudomallei*. Other examples of bacterial species that are actually clones with distinctive biology and ecology include *Bacillus anthracis*[120] and *Salmonella typhi*.[121] *Neisseria gonorrhoeae* strains form a

Downloaded From Vet-eBooks.com

tight cluster at the end of a long branch arising from the meningococcal cluster,[122] supporting the hypothesis that gonococci arose relatively recently as a strain of human pharyngeal *Neisseria* species that acquired the ability to colonize the genital tract and be transmitted by the sexual route.[123] MLST has also proven useful in the context of taxonomic groups with low genomic representation, for example, neglected diseases or industrial microbes,[124,125] and on studies where large numbers of samples are analyzed.[126,127] For instance, Nuñez et al. interrogated the genetic structure of the bioleaching microbe *Acidithiobacillus caldus* and found overall low genetic diversity from different geographic locations, which supports current taxonomic assignments and suggests that bioprocesses constrain genetic diversity.[125]

7. Conclusions and Prospects

MLST has become a standard and flexible approach for characterizing bacteria and some eukaryotes mainly due to the existence of comprehensive databases and its broad implementation in clinical laboratory settings, from basic research laboratories (PCR + Sanger) to core sequencing facilities (cgMLST; in silico MLST). MLST has expanded its basic scheme to incorporate more and new molecular markers, such as ribosomal proteins and large matrices of orthologous genes (gene-by-gene approach), and more recently, to integrate pan and core-genome concepts as well as draft and full genomes. Two-tier strategies currently being applied to human microbiome research where investigations start by using MLST to type as many samples as possible, and continue by delving further into isolate groups of particular interest by using whole-genome sequencing are already in practice.[117,128]

New MLST-genome strategies will also provide more accurate and robust estimates of population genetic parameters under more complex and realistic statistical models such as those based on the coalescent model.[129] Moreover, within this framework, epidemiological data can also be integrated; hence more comprehensive and faster assessments of pathogen dynamics can be achieved. Microbial genomics is expanding outside research laboratories into clinical practice and molecular diagnostics.[130,131] One can only assume that classical or expanded forms of MLST will remain a key component of the microbial genomicist's toolkit toward understanding the ecology and evolution of infectious diseases.

Acknowledgments

M.P.-L. was funded by a DC D-CFAR Research Award from the District of Columbia Developmental Center for AIDS Research (P30AI087714), by a University Facilitating Fund award from George Washington University, and a K12 Career Development Program 5 K12 HL119994 award. M.A. was supported by the Portuguese Government through the FCT Starting Grant IF/00955/2014. E.C.N. was funded by "CONICYT + PAI/CONCURSO NACIONAL APOYO AL RETORNO DE INVESTIGADORES/AS DESDE EL EXTRANJERO, CONVOCATORIA 2014 + FOLIO 82140008."

Downloaded From Vet-eBooks.com

References

1. Cooper JE, Feil EJ. Multilocus sequence typing—what is resolved? *Trends Microbiol* 2004;**12**:373–7.
2. Foley SL, Lynne AM, Nayak R. Molecular typing methodologies for microbial source tracking and epidemiological investigations of gram-negative bacterial foodborne pathogens. *Infect Genet Evol* 2009;**9**:430–40.
3. Maiden MC, Bygraves JA, Feil E, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* 1998;**95**:3140–5.
4. Maiden MC. Multilocus sequence typing of bacteria. *Annu Rev Microbiol* 2006;**60**: 561–88.
5. Pérez-Losada M, Cabezas P, Castro-Nallar E, Crandall KA. Pathogen typing in the genomics era: MLST and the future of molecular epidemiology. *Infect Genet Evol* 2013; **16**:38–53.
6. Jolley KA, Maiden MC. Using MLST to study bacterial variation: prospects in the genomic era. *Future Microbiol* 2014;**9**:623–30.
7. Maiden MC, van Rensburg MJJ, Bray JE, et al. MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat Rev Microbiol* 2013;**11**:728–36.
8. Urwin R, Maiden MC. Multi-locus sequence typing: a tool for global epidemiology. *Trends Microbiol* 2003;**11**:479–87.
9. Sullivan CB, Diggle MA, Clarke SC. Multilocus sequence typing: data analysis in clinical microbiology and public health. *Mol Biotechnol* 2005;**29**:245–54.
10. Boers SA, van der Reijden WA, Jansen R. High-throughput multilocus sequence typing: bringing molecular typing to the next level. *PLoS One* 2012;**7**:e39630.
11. Jolley KA, Bliss CM, Bennett JS, et al. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology* 2012;**158**: 1005–15.
12. Larsen MV, Cosentino S, Rasmussen S, et al. Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol* 2012;**50**:1355–61.
13. Pérez-Losada M, Browne EB, Madsen A, Wirth T, Viscidi RP, Crandall KA. Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data. *Infect Genet Evol* 2006;**6**:97–112.
14. Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005;**437**:376–80.
15. Chen Y, Frazzitta AE, Litvintseva AP, et al. Next generation multilocus sequence typing (NGMLST) and the analytical software program MLSTEZ enable efficient, cost-effective, high-throughput, multilocus sequencing typing. *Fungal Genet Biol* 2015;**75**: 64–71.
16. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 2013;**79**:5112–20.
17. O'Halloran DM. PrimerMapper: high throughput primer design and graphical assembly for PCR and SNP detection. *Sci Rep* 2016;**6**:20631.
18. Untergasser A, Cutcutache I, Koressaar T, et al. Primer3—new capabilities and interfaces. *Nucleic Acids Res* 2012;**40**:e115.
19. Pérez-Losada M, Crandall KA, Zenilman J, Viscidi RP. Temporal trends in gonococcal population genetics in a high prevalence urban community. *Infect Genet Evol* 2007;**7**: 271–8.

Downloaded From Vet-eBooks.com

20. Carattoli A, Zankari E, García-Fernández A, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 2014;**58**:3895−903.
21. Yoshida C, Kruczkiewicz P, Laing C, et al. The *Salmonella In Silico* Typing Resource (SISTR): an open web-accessible tool for rapidly typing and subtyping draft *Salmonella* genome assemblies. *PLoS One* 2015;**11**:e0147101.
22. Kruczkiewicz P, Mutschall S, Barker D, et al. MIST: a tool for rapid in silico generation of molecular data from bacterial genome sequences. *Bioinformatics* 2013:316−23.
23. Inouye M, Dashnow H, Raven L-A, et al. SRST2: rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med* 2014;**6**:1−16.
24. Cookson BD, Robinson DA, Monk AB, et al. Evaluation of molecular typing methods in characterizing a European collection of epidemic methicillin-resistant *Staphylococcus aureus* strains: the HARMONY collection. *J Clin Microbiol* 2007;**45**:1830−7.
25. Metzgar D, Baynes D, Hansen CJ, et al. Inference of antibiotic resistance and virulence among diverse group A *Streptococcus* strains using *emm* sequencing and multilocus genotyping methods. *PLoS One* 2009;**4**:e6897.
26. Siarkou VI, Vorimore F, Vicari N, et al. Diversification and distribution of Ruminant *Chlamydia abortus* clones assessed by mlst and MLVA. *PLoS One* 2015;**10**:e0126433.
27. van Cuyck H, Pichon B, Leroy P, et al. Multiple-locus variable-number tandem-repeat analysis of *Streptococcus pneumoniae* and comparison with multiple loci sequence typing. *BMC Microbiol* 2012;**12**:241.
28. Marsh JW, O'Leary MM, Shutt KA, et al. Multilocus variable-number tandem-repeat analysis and multilocus sequence typing reveal genetic relationships among *Clostridium difficile* isolates genotyped by restriction endonuclease analysis. *J Clin Microbiol* 2010;**48**: 412−8.
29. de Been M, Pinholt M, Top J, et al. A core genome MLST scheme for high-resolution typing of *Enterococcus faecium*. *J Clin Microbiol* 2015.
30. Hall BG, Ehrlich GD, Hu FZ. Pan-genome analysis provides much higher strain typing resolution than multi-locus sequence typing. *Microbiology* 2010;**156**:1060−8.
31. Leekitcharoenphon P, Lukjancenko O, Friis C, Aarestrup FM, Ussery DW. Genomic variation in *Salmonella enterica* core genes for epidemiological typing. *BMC Genomics* 2012;**13**:88.
32. Chan MS, Maiden MC, Spratt BG. Database-driven multi locus sequence typing (MLST) of bacterial pathogens. *Bioinformatics* 2001;**17**:1077−83.
33. Jolley KA, Chan MS, Maiden MC. mlstdbNet − distributed multi-locus sequence typing (MLST) databases. *BMC Bioinformatics* 2004;**5**:86.
34. Jolley KA, Maiden MC. BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 2010;**11**:595.
35. Jolley KA, Maiden MC. AgdbNet − antigen sequence database software for bacterial typing. *BMC Bioinformatics* 2006;**7**:314.
36. Jefferies J, Clarke SC, Diggle MA, Smith A, Dowson C, Mitchell T. Automated pneumococcal MLST using liquid-handling robotics and a capillary DNA sequencer. *Mol Biotechnol* 2003;**24**:303−7.
37. Parkhill J, Sebaihia M, Preston A, et al. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat Genet* 2003;**35**:32−40.
38. Enright MC, Spratt BG. Multilocus sequence typing. *Trends Microbiol* 1999;**7**:482−7.
39. Loubna T, Pérez-Losada M, Gu W, et al. Population dynamics of *Neisseria gonorrhoeae* in Shanghai, China: a comparative study. *BMC Infect Dis* 2010;**10**:13.

Downloaded From Vet-eBooks.com

40. Jolley KA, Feil EJ, Chan MS, Maiden MC. Sequence type analysis and recombinational tests (START). *Bioinformatics* 2001;**17**:1230−1.
41. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol* 2004;**186**:1518−30.
42. Francisco AP, Bugalho M, Ramirez M, Carrico JA. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinformatics* 2009;**10**: 152.
43. Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M, Carrio JA. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics* 2012;**13**:87.
44. Wiens JJ. Reconstructing phylogenies from allozyme data: comparing method performance with congruence. *Biol J Linn Soc* 2000;**70**:613−32.
45. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;**30**:772−80.
46. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;**32**:1792−7.
47. Lemmon AR, Moriarty EC. The importance of proper model assumption in bayesian phylogenetics. *Syst Biol* 2004;**53**:265−77.
48. Posada D, Buckley TR. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and Bayesian approaches over likelihood ratio tests. *Syst Biol* 2004;**53**:793−808.
49. Arenas M. Trends in substitution models of molecular evolution. *Front Genet* 2015;**6**:319.
50. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 2012;**9**:772.
51. Luo A, Qiao H, Zhang Y, et al. Performance of criteria for selecting evolutionary models in phylogenetics: a comprehensive study based on simulated datasets. *BMC Evol Biol* 2010; **10**:242.
52. Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* 1981;**17**:368−76.
53. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 2006;**22**:2688−90.
54. Stamatakis A, Aberer AJ, Goll C, Smith SA, Berger SA, Izquierdo-Carrasco F. RAxML-Light: a tool for computing terabyte phylogenies. *Bioinformatics* 2012;**28**:2064−6.
55. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985;**39**:783−91.
56. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;**59**:307−21.
57. Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 2001;**294**:2310−4.
58. Ronquist F, Teslenko M, van der Mark P, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 2012;**61**:539−42.
59. Rambaut A, Drummond AJ. *Tracer: MCMC trace analysis tool*. 1.5 ed. Edinburgh: Institute of Evolutionary Biology; 2009. http://tree.bio.ed.ac.uk/software/tracer/.
60. Zoller S, Boskova V, Anisimova M. Maximum-likelihood tree estimation using codon substitution models with multiple partitions. *Mol Biol Evol* 2015;**32**:2208−16.
61. Arenas M, Posada D. Coalescent simulation of intracodon recombination. *Genetics* 2010; **184**:429−37.

Downloaded From Vet-eBooks.com

62. Shimodaira H, Hasegawa M. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 2001;**17**:1246−7.
63. Liu L. BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics* 2008;**24**:2542−3.
64. Heled J, Drummond AJ. Bayesian inference of species trees from multilocus data. *Mol Biol Evol* 2010;**27**:570−80.
65. Schierup MH, Hein J. Consequences of recombination on traditional phylogenetic analysis. *Genetics* 2000;**156**:879−91.
66. Posada D, Crandall KA. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci USA* 2001;**98**:13757−62.
67. Woolley SW, Posada D, Crandall KA. A comparison of phylogenetic network methods using computer simulation. *PLoS Comput Biol* 2008;**3**:e1913.
68. Cassens I, Mardulyn P, Milinkovitch MC. Evaluating intraspecific "network" construction methods using simulated sequence data: do existing algorithms outperform the global maximum parsimony approach? *Syst Biol* 2005;**54**:363−72.
69. Templeton AR, Crandall KA, Sing CF. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 1992;**132**:619−33.
70. Huson DH. SplitsTree: a program for analyzing and visualizing evolutionary data. *Bioinformatics* 1998;**14**:68−73.
71. Didelot X, Falush D. Inference of bacterial microevolution using multilocus sequence data. *Genetics* 2007;**175**:1251−66.
72. Martin DP, Lemey P, Posada D. Analysing recombination in nucleotide sequences. *Mol Ecol Resour* 2011;**11**:943−55.
73. Posada D, Crandall KA, Holmes EC. Recombination in evolutionary genomics. *Annu Rev Genet* 2002;**36**:75−97.
74. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 2010;**26**:2462−3.
75. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. GARD: a genetic algorithm for recombination detection. *Bioinformatics* 2006;**22**:3096−8.
76. Arenas M, Lopes JS, Beaumont MA, Posada D. CodABC: a computational framework to coestimate recombination, substitution, and molecular adaptation rates by approximate Bayesian computation. *Mol Biol Evol* 2015;**32**:1109−12.
77. Pérez-Losada M, Porter ML, Tazi L, Crandall KA. New methods for inferring population dynamics from microbial sequences. *Infect Genet Evol* 2007;**7**:24−43.
78. Pearse DE, Crandall K. Beyond F_{ST}: analysis of population genetic data for conservation. *Conserv Genet* 2004;**5**:585−602.
79. Excoffier L, Heckel G. Computer programs for population genetics data analysis: a survival guide. *Nat Rev Genet* 2006;**7**:745−58.
80. Waples RS, Gaggiotti O. What is a population? an empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Mol Ecol* 2006;**15**:1419−39.
81. Bashalkhanov S, Pandey M, Rajora OP. A simple method for estimating genetic diversity in large populations from finite sample sizes. *BMC Genet* 2009;**10**:84.
82. Minin VN, Bloomquist EW, Suchard MA. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol Biol Evol* 2008;**25**: 1459−71.
83. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 2012;**29**:1969−73.

Downloaded From Vet-eBooks.com

84. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 2007;**24**: 1586–91.
85. Yang Z, Wong WS, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 2005;**22**:1107–18.
86. Kosakovsky Pond SL, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 2005;**21**:676–9.
87. Lemey P, Rambaut A, Welch JJ, Suchard MA. Phylogeography takes a relaxed random walk in continuous space and time. *Mol Biol Evol* 2010;**27**:1877–85.
88. Bielejec F, Rambaut A, Suchard MA, Lemey P. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics* 2011;**27**:2910–2.
89. McAdam PR, Templeton KE, Edwards GF, et al. Molecular tracing of the emergence, adaptation, and transmission of hospital-associated methicillin-resistant *Staphylococcus aureus*. *Proc Natl Acad Sci USA* 2012;**109**:9107–12.
90. Weinert LA, Welch JJ, Suchard MA, Lemey P, Rambaut A, Fitzgerald JR. Molecular dating of human-to-bovid host jumps by *Staphylococcus aureus* reveals an association with the spread of domestication. *Biol Lett* 2012;**8**:829–32.
91. Palazzo IC, Pitondo-Silva A, Levy CE, da Costa Darini AL. Changes in vancomycin-resistant *Enterococcus faecium* causing outbreaks in Brazil. *J Hosp Infect* 2011;**79**:70–4.
92. Vanderkooi OG, Church DL, MacDonald J, Zucol F, Kellner JD. Community-based outbreaks in vulnerable populations of invasive infections caused by *Streptococcus pneumoniae* serotypes 5 and 8 in Calgary, Canada. *PLoS One* 2011;**6**:e28547.
93. Haran KP, Godden SM, Boxrud D, Jawahir S, Bender JB, Sreevatsan S. Prevalence and characterization of *Staphylococcus aureus*, including methicillin-resistant *Staphylococcus aureus*, isolated from bulk tank milk from Minnesota dairy farms. *J Clin Microbiol* 2012; **50**:688–95.
94. Ibarz-Pavon AB, Morais L, Sigauque B, et al. Epidemiology, molecular characterization and antibiotic resistance of *Neisseria meningitidis* from patients ≤15 years in Manhica, rural Mozambique. *PLoS One* 2011;**6**:e19717.
95. Martin V, Maldonado-Barragan A, Moles L, et al. Sharing of bacterial strains between breast milk and infant feces. *J Hum Lact* 2012;**28**:36–44.
96. Walker AS, Eyre DW, Wyllie DH, et al. Characterisation of *Clostridium difficile* hospital ward-based transmission using extensive epidemiological data and molecular typing. *PLoS Med* 2012;**9**:e1001172.
97. Egger R, Korczak BM, Niederer L, Overesch G, Kuhnert P. Genotypes and antibiotic resistance of *Campylobacter coli* in fattening pigs. *Vet Microbiol* 2012;**155**:272–8.
98. Romano S, Aujoulat F, Jumas-Bilak E, et al. Multilocus sequence typing supports the hypothesis that *Ochrobactrum anthropi* displays a human-associated subpopulation. *BMC Microbiol* 2009;**9**.
99. Oteo J, Cuevas O, Lopez-Rodriguez I, et al. Emergence of CTX-M-15-producing *Klebsiella pneumoniae* of multilocus sequence types 1, 11, 14, 17, 20, 35 and 36 as pathogens and colonizers in newborns and adults. *J Antimicrob Chemother* 2009;**64**:524–8.
100. Lee MY, Choi HJ, Choi JY, et al. Dissemination of ST131 and ST393 community-onset, ciprofloxacin-resistant *Escherichia coli* clones causing urinary tract infections in Korea. *J Infect* 2010;**60**:146–53.
101. Yazdankhah SP, Kriz P, Tzanakaki G, et al. Distribution of serogroups and genotypes among disease-associated and carried isolates of *Neisseria meningitidis* from the Czech Republic, Greece, and Norway. *J Clin Microbiol* 2004;**42**:5146–53.

Downloaded From Vet-eBooks.com

102. Ruiz-Garbajosa P, Bonten MJ, Robinson DA, et al. Multilocus sequence typing scheme for *Enterococcus faecalis* reveals hospital-adapted genetic complexes in a background of high rates of recombination. *J Clin Microbiol* 2006;**44**:2220−8.
103. Leavis HL, Bonten MJ, Willems RJ. Identification of high-risk enterococcal clonal complexes: global dispersion and antibiotic resistance. *Curr Opin Microbiol* 2006;**9**: 454−60.
104. Enright MC, Spratt BG. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. *Microbiology* 1998;**144**(Pt 11):3049−60.
105. Feil EJ, Cooper JE, Grundmann H, et al. How clonal is *Staphylococcus aureus*? *J Bacteriol* 2003;**185**:3307−16.
106. Kalia A, Spratt BG, Enright MC, Bessen DE. Influence of recombination and niche separation on the population genetic structure of the pathogen *Streptococcus pyogenes*. *Infect Immun* 2002;**70**:1971−83.
107. Robinson DA, Enright MC. Evolutionary models of the emergence of methicillin-resistant *Staphylococcus aureus*. *Antimicrob Agents Chemother* 2003;**47**:3926−34.
108. Choudhury B, Risley CL, Ghani AC, et al. Identification of individuals with gonorrhoea within sexual networks: a population-based study. *Lancet* 2006;**368**:139−46.
109. Stadler T, Kouyos R, von Wyl V, et al. Estimating the basic reproductive number from viral sequence data. *Mol Biol Evol* 2012;**29**:347−57.
110. Popinga A, Vaughan T, Stadler T, Drummond AJ. Inferring epidemiological dynamics with Bayesian coalescent inference: the merits of deterministic and stochastic models. *Genetics* 2015;**199**:595−607.
111. du Plessis L, Stadler T. Getting to the root of epidemic spread with phylodynamic analysis of genomic data. *Trends Microbiol* 2015;**23**:383−6.
112. Volz EM, Frost SD. Inferring the source of transmission with phylogenetic data. *PLoS Comput Biol* 2013;**9**:e1003397.
113. Rasmussen DA, Volz EM, Koelle K. Phylodynamic inference for structured epidemiological models. *PLoS Comput Biol* 2014;**10**:e1003570.
114. Hamby SE, Joseph S, Forsythe SJ, Chuzhanova N. In silico identification of pathogenic strains of *Cronobacter* from biochemical data reveals association of inositol fermentation with pathogenicity. *BMC Microbiol* 2011;**11**:1.
115. Stasiewicz MJ, Oliver HF, Wiedmann M, den Bakker HC. Whole-genome sequencing allows for improved identification of persistent *Listeria monocytogenes* in food-associated environments. *Appl Environ Microbiol* 2015;**81**:6024−37.
116. Wong VK, Baker S, Pickard DJ, et al. Phylogeographical analysis of the dominant multidrug-resistant H58 clade of *Salmonella* Typhi identifies inter-and intracontinental transmission events. *Nat Genet* 2015;**47**:632−9.
117. Holt KE, Wertheim H, Zadoks RN, et al. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. *Proc Natl Acad Sci* 2015;**112**:E3574−81.
118. Fraser C, Hanage WP, Spratt BG. Recombination and the nature of bacterial speciation. *Science* 2007;**315**:476−80.
119. Godoy D, Randle G, Simpson AJ, et al. Multilocus sequence typing and evolutionary relationships among the causative agents of melioidosis and glanders, *Burkholderia pseudomallei* and *Burkholderia mallei*. *J Clin Microbiol* 2003;**41**:2068−79.
120. Priest FG, Barker M, Baillie LW, Holmes EC, Maiden MC. Population structure and evolution of the *Bacillus cereus* group. *J Bacteriol* 2004;**186**:7959−70.

Downloaded From Vet-eBooks.com

121. Kidgell C, Reichard U, Wain J, et al. *Salmonella typhi*, the causative agent of typhoid fever, is approximately 50,000 years old. *Infect Genet Evol* 2002;**2**:39−45.
122. Hanage WP, Fraser C, Spratt BG. Sequences, sequence clusters and bacterial species. *Philos Trans R Soc Lond B Biol Sci* 2006;**361**:1917−27.
123. Vazquez JA, de la Fuente L, Berron S, et al. Ecological separation and genetic isolation of *Neisseria gonorrhoeae* and *Neisseria meningitidis*. *Curr Biol* 1993;**3**:567−72.
124. Boonsilp S, Thaipadungpanit J, Amornchai P, et al. A single multilocus sequence typing (MLST) scheme for seven pathogenic *Leptospira* species. *PLoS Negl Trop Dis* 2013;**7**: e1954.
125. Nuñez H, Loyola D, Cárdenas JP, Holmes DS, Johnson DB, Quatrini R. Multilocus sequence typing scheme for *Acidithiobacillus caldus* strain evaluation and differentiation. *Res Microbiol* 2014;**165**:735−42.
126. Jacquot M, Bisseux M, Abrial D, et al. High-throughput sequence typing reveals genetic differentiation and host specialization among populations of the *Borrelia burgdorferi* species complex that infect rodents. *PLoS One* 2014;**9**:e88581.
127. Rosales R, Churchward C, Schnee C, et al. Global multilocus sequence typing analysis of *Mycoplasma bovis* isolates reveals two main population clusters. *J Clin Microbiol* 2015; **53**:789−94.
128. Franzosa EA, Hsu T, Sirota-Madi A, et al. Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nat Rev Microbiol* 2015;**13**:360−72.
129. Mather AE, Vaughan TG, French NP. Molecular approaches to understanding transmission and source attribution in nontyphoidal *Salmonella* and their application in Africa. *Clin Infect Dis* 2015;**61**:S259−65.
130. Gardy J, Loman NJ, Rambaut A. Real-time digital pathogen surveillance—the time is now. *Genome Biol* 2015;**16**.
131. Luheshi LM, Raza S, Peacock SJ. Moving pathogen genomics out of the lab and into the clinic: what will it take? *Genome Med* 2015;**7**.

Downloaded From Vet-eBooks.com

Next-Generation Sequencing, Bioinformatics, and Infectious Diseases

17

R. van Aerle [1], *M. van der Giezen* [2]
[1]Centre for Environment, Fisheries and Aquaculture Science, Weymouth, United Kingdom;
[2]University of Exeter, Exeter, United Kingdom

1. Analyzing Big Data

Bioinformatics has been around for nearly 40 years and started as the study of informatics processes in biotic systems.[1,2] As most of biology and medical sciences is becoming more and more "big data", the introduction of bioinformatics in almost all subdisciplines has led to multiple interpretations of what bioinformatics actually entails. It is clear that bioinformatics has played an important role in analyzing the vast expanse of sequenced genomes and in particular in meaningful comparative analyses of many large datasets. It is simply no longer feasible to analyze such datasets by hand. However, with immense computational power at hand, there is a danger of overinterpreting data, and a good biological understanding of the system under study is still essential.

At the heart of bioinformatics is the generation of vast amounts of sequence data. It is clear that the development of ever more efficient and capable sequencing platforms resulted in the production of many sequences. Only recently, data from newer sequencing platforms has overtaken "standard" GenBank submission (see Fig. 17.1). Bioinformatics is tasked with making sense of this data, mining it, storing it, and disseminating it, and ensuring valid biological conclusions can be made.

Many fields have benefitted from these developments. Genomics has obviously been at the forefront but also large-scale transcriptomics studies have allowed a glimpse into the inner workings of cells under different growth conditions. Evolution, epidemiology, and ecology are all fields that require bioinformatics to accurately process and place the various sources of data into context.

In this chapter, we aim to discuss relevant examples from infectious disease research that have benefitted from the next-generation sequencing surge and how bioinformatics enabled the dissemination of this information.

2. Comparative Genomics

Genome sequencing is now taken for granted as being a rather routine procedure. However, this has not always been the case and the first genomes were scientific achievements of the first order. These first-sequenced genomes were those from

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00017-2
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

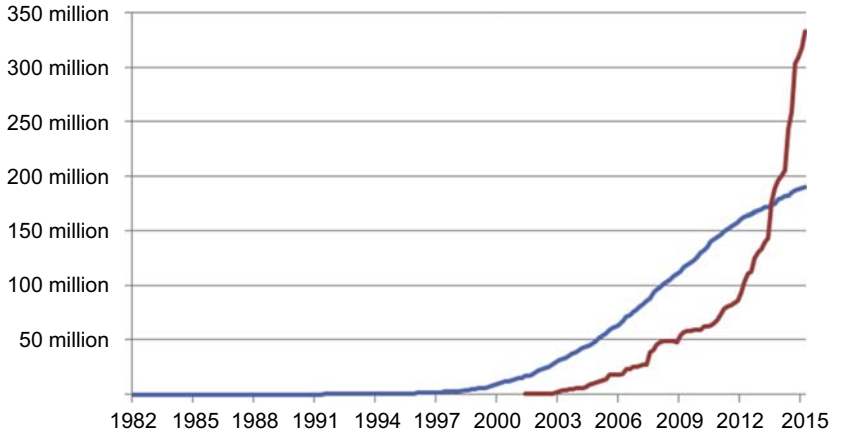


Figure 17.1 The growth of sequence submissions to GenBank. In June 2014, the number of whole-genome sequences (WGS) overtook the number individual submitted sequenced. The current growth and low price for whole genome sequences likely means that the WGS submissions keep increasing dramatically. For further information see http://www.ncbi.nlm.nih.gov/genbank/statistics/.

several bacteriophages such as MS2[3] and φX174[4] and the Simian virus 40 genome.[5] These were relatively small and any larger genomes were at the time hampered due to considerable technical challenges. After Applied Biosystems produced the first automated sequencers in the late 1980s, biologists could finally tackle "real" genomes. Obviously, the first bacterial genomes that were completed were those of pathogens. The bacterial pathogens *Haemophilus influenza*,[6] *Mycoplasma genitalium*,[7] and *Helicobacter pylori*[8] were sequenced in close succession. In addition to pathogens, the genomes of model organisms such as *Saccharomyces cerevisiae*[9] and *Caenorhabditis elegans*[10] were completed early on as well. When we write this chapter (Spring 2016), Genbank lists over 66,000 prokaryotic genomes and nearly 3000 eukaryotic genomes, so it is clear that sequencing has taken off considerably since mid-1990s. Nowadays it is clearly the case that "what you do with it" is more important than simply having produced a genome and it certainly will not result in a publication in scientific journals, such as Nature or Science, as the early genomes easily did. The following examples are infectious disease genome projects where clear hypotheses were formulated utilizing vast amounts of sequence data.

Malaria is still a major killer and although numbers of deaths have come down considerably in the last decade, it still claims the lives of nearly half a million people annually.[11] The main causative agent is the apicomplexan parasite *Plasmodium falciparum* and it is now clear that this parasite originated in the African apes.[12] However, comparative analyses were mainly limited to mitochondrial genome sequences. Although the *P. falciparum* genome has been available for over a decade,[13] only recently the high-quality genome of a chimpanzee-infecting *Plasmodium* species, *Plasmodium reichenowi*, became available.[14] Both genomes are near identical in

Downloaded From Vet-eBooks.com

size and gene content. When the subtelomeric regions are ignored, both genomes demonstrate a complete genomic synteny, clearly suggestive of their recent shared ancestry. Both *Plasmodium* species share about 5000 genes. Three genes that are present in *P. reichenowi* are missing in *P. falciparum* and only one gene in the latter is not present in the *P. reichenowi* genome. The human parasite contains 16 genes that are now pseudogenes in *P. reichenowi* and this is the case for 19 genes in the opposite direction. It is clear that these genomes are indeed very similar. Nonetheless, one is a chimpanzee parasite while the other is a human pathogen. Host-specific genes in *P. falciparum* are known to reside in the subtelomeric region of the genome. Many of these genes are exported to the host erythrocytes and it would be expected that this might be a region where the chimpanzee parasite differs from the human one. However, almost all *P. falciparum* genes could be assigned an ortholog in the *P. reichenowi* genome while only a few on the latter genome were pseudogenes. There were differences in the sizes of two multigene families that encode proteins that are presented on the surface of infected erythrocytes. The *rif* and *stevor* gene families are present in all *Plasmodium* species despite the fact that their function is not known.[15] The chimpanzee parasite contains close to 600 *rif* genes and 66 *stevor* genes while the human *Plasmodium* species has 185 and 42, respectively. So, although these genes clearly play some role in the host−pathogen interface, the human parasite manages to maintain a successful infection despite a reduced gene set in these two multigene families. When analyzing both genomes for selective pressures to retain genes, 77% of the top 100 most divergent genes had no known function compared to 38% of genes on the whole *P. falciparum* genome. Of these 100 genes, 31 contained motifs suggesting a role in erythrocyte import. The genes that are missing or that are pseudogenes in *P. reichenowi* are present in this 100-gene set of most divergent genes and all are members of gene families involved in erythrocyte invasion. Hence, it is clear that many of the genes identified to be different between the chimpanzee and human *Plasmodium* species are involved in red blood cell invasion. This genome comparison identified some clear candidate genes for further studies and might shed additional light on the crucial interactions that occur at the host−pathogen interface.[14]

Many infectious agents are transmitted by ticks; in fact, ticks are transmitting a larger variety of disease agents than any other blood-feeding arthropod.[16] Tick-borne diseases, such as the notorious Lyme disease, result in thousands of animal and human deaths each year. However, ticks also transmit other diseases such as babesiosis and granulocytic anaplasmosis. The 2016 completed genome of the tick *Ixodes scapularis* provided insight into the biology and adaptations of this very successful parasite of mammals.[16] During a blood meal, ticks can increase near 100-fold in body weight while during periods off their host, they are able to survive without feeding for several months.[16] As ticks feed for a period lasting many days, they need to be able to securely attach themselves to their hosts while fending off any attempts by the host to attack the parasite. When comparing salivary gland products, it is clear that ticks have evolved mechanisms to enhance their survival. The tick genome encodes the largest repertoire of Kunitz domain[17] containing proteins. These domains are involved in protease inhibition, inhibiting coagulation, angiogenesis, and vasodilation. The tick genome contains 74 genes in this class compared to less than

Downloaded From Vet-eBooks.com

nine in the medically important mosquito vectors, such as *Aedes*, *Culex*, and *Anopheles*.[16] Ticks are not able to produce heme, a likely consequence of their ability to scavenge heme from their hosts. However, some heme biosynthetic genes were identified on the tick genome, but these may be remnants of a once-complete pathway.[18] More interesting is their mode of hemoglobin digestion that results in total degradation into dipeptides and amino acids. Ticks employ similar enzymes as found in other hemoglobinolytic organisms. *Cathepsin D* and *L* and serine carboxypeptidases are all present in multiple copies[16] suggesting their importance for the organism. The unique hemoglobin degradation involving specialized organelles (hemosomes) might therefore be better targets when developing antitick drugs. Especially the special enzymes involved in heme storage during the long periods between feeding and egg development, such as hemelipoglyco-carrier proteins and vitellogenins, might be interesting in this respect. Although blood feeding is an effective strategy to obtain nutrient, the blood of hosts might contain many dangers for the parasite. The tick genome contains 206 CYP450 genes, which is a record of possible detoxifying CYP genes for any organism.[16] In addition, the tick immune system contains several strategies to prevent pathogen invasion. The Toll system, several other immune pathways, and RNAi were found in the genome. Nonetheless, pathogens such as the Lyme disease−transmitting *Borrelia burgdorferi* manage to evade these systems and maintain a presence within the tick. Ticks spend extended periods off host and their survival is crucially dependent on their ability to sense the proximity of new host. *Ixodes* contains 62 gustatory receptors to aid in host sensing. These genes fall in three clades, one of them contains 42 genes and this clade is unique to ticks.[16] Interestingly, the tick genome project also included the complete genome of a bacterial endosymbiont, a *Rickettsia* species specific to ticks.

A 2016 international collaborative effort was initiated to investigate the Ebola virus disease outbreak in West Africa.[19] The team used mobile laboratories and portable nanopore sequencers to sequence and analyze Ebola.[19] They used 11−19 primer pairs to generate amplicons that covered >97% of the Ebola genome and these were subsequently sequenced using the MinION (Oxford Nanopore Technologies). The genome sequencing workflow, including amplification, sequence library preparation, and sequencing could be performed within 1−2 days. Within the first 10 days of real-time sequencing, they concluded that the persisting Guinean Ebola cases belonged to two major lineages (GN1 and SL3). Lineage GN1 was found to be largely confined to Guinea, whereas previous research indicated that lineage SL3 was derived from Sierra Leone in 2014 and moved to Guinea toward the end of that year.[20] Integration of these MinION sequence datasets with other previously published data suggested that both lineages had also been seen in Sierra Leone early in 2015, suggesting transmission between the countries. A dedicated website showing real-time analysis of Ebola evolution was established (http://ebola.nextstrain.org/) and updated throughout the study to provide the latest developments in the evolution of the virus. This real-time genomic surveillance study demonstrates the state-of-the-art in real-time, long-read sequencing of highly infectious global pathogens and could serve as a basis for responses to future outbreaks. Indeed, this framework is currently being used to study the Zika virus outbreak that is currently ongoing in Latin America.

Downloaded From Vet-eBooks.com

3. Transcriptomics

Having access to an organism's genome can be a valuable resource informing further studies. However, not all genes will be expressed at any given moment, and the actual environment plays an important role in which genes are being expressed at any one time. Gene expression analysis (transcriptomics) can be used to study the responses of an organism to a given environment or, for example, during host−pathogen interactions. By investigating changes in gene expression patterns during an infection process or over the life cycle of pathogens, key molecular processes/pathways or individual (virulence) genes can be discovered. Transcripts (mRNA, but also noncoding RNAs) from both the host and pathogen can be analyzed simultaneously, potentially shedding light on important processes leading to disease.[21]

The causative agent of Denman Island disease, a notifiable disease of commercially important oysters, is the parasite *Mikrocytos mackini*.[22] *M. mackini* is a nondescript organism known as a microcell. Its taxonomic placement has been unclear and the standard molecular marker for taxonomic classification, the small subunit ribosomal RNA gene, is highly divergent in this organism.[23] Burki et al. embarked on a large-scale RNA-Seq approach to shed light on the taxonomic placement of this commercially important pathogen and also to understand the cell biology of this rather primitive-looking organism.[24] As an environmental sequencing approach is fraught with danger, the authors carefully removed any sequences that appeared to be host derived or possibly prokaryotic in origin. This left over 15,000 possible *M. mackini*−specific sequences. From this set, they identified over 250 that had been used in other large-scale phylogenetic studies. Single-gene phylogenies were all poorly resolved and hinted at an accelerated rate of evolution for *Mikrocytos* genes. Bioinformatics analyses of these genes resulted in just over 100 that seemed suitable for further analyses. Irrespectively, initial results suggested a relationship with the rhizarian supergroup[25] that resulted in the authors adding 16 rhizarian-specific genes to the dataset. The final dataset to identify the taxonomic placement of *M. mackini* contained 119 protein-encoding genes and over 23,000 amino acid positions.[24] This large-scale concatenated phylogenetic analysis clearly placed this parasite as a member of the Rhizaria and also as one of the most fast-evolving eukaryotes known.[24] *Mikrocytos*' simple cell biological features[22] rekindled questions about its mitochondrial status.[26] In order to address this, the authors analyzed all 15,000 or so genes using a mitochondrial-prediction pipeline.[27] This resulted in 88 putative mitochondrial proteins (instead of the ∼1000 found in typical mitochondriate eukaryotes) and after manual curation, only four were found to be genuine mitochondrial proteins with apparent amino-terminal mitochondrial targeting signals.[24] As all four identified proteins are involved in the genuine mitochondrial iron sulfur cluster assembly pathway,[28] it seems that *M. mackini* does have mitochondria of some sort despite earlier cell biological studies.[22] Although localization studies remain to be conducted to confirm the presence of a mitochondrion, this study clearly shows how directed transcriptomics studies can inform further cell biological studies. The large-scale phylogenetic study also aids attempts to identify possible drug targets as its inclusion within the Rhizaria, which includes other oyster

Downloaded From Vet-eBooks.com

pathogens, such as *Haplosporidium*, might direct a more focused approach to prevent or treat this notifiable disease.

Another strength of transcriptomics is the possible comparative approach understanding the transcriptomic response to different environmental stimuli or during developmental stages. The human intestinal parasite *Entamoeba histolytica* is the third most leading cause of protistan death and responsible for up to 100,000 deaths annually.[29] *Entamoeba* has a biphasic lifestyle alternating between a motile feeding stage called trophozoite and an immotile survival stage called cyst. Cysts are produced in the lower parts of the intestinal tract and allow *Entamoeba* to survive the environmental conditions outside its host. Poor sanitary conditions ultimately lead to the uptake of cysts via contaminated drinking water and thereby completing the life cycle of this parasite. The actual cues that induce cyst formation are not clear and more frustratingly, it is currently not possible to induce cyst formation for *E. histolytica.* For this reason, the lizard pathogenic species *Entameoba invadens* is used to study cyst formation as it can be induced in vitro in the laboratory.[30] In order to understand the underlying molecular events that lead to the production of cysts, Ehrenkaufer et al. conducted a comparative transcriptomics during encystation using next-generation SOLiD sequencing.[31] Their paper also reported a Sanger genome for *E. invadens* with an only fourfold average coverage. Although poor coverage, most main features could be compared with the genome of the human pathogen *E. histolytica*.[32] The high-quality SOLiD data was more interesting especially as the authors took samples from various time points during cyst formation. RNA was extracted at 0, 8, 24, 48, and 72 h post induction. There were two biological replicates but as the cyst formation process cannot be synchronized, the authors were forced to investigate major gene regulation patterns. As many years of *Entamoeba* research had produced a good, but sparse, list of genes known to be involved in cyst formation, the authors could benchmark their more recent high-throughput data. The over 11,000 putative *E. invadens* genes were followed during cyst formation, and up- and downregulations were assessed. Nearly 10,000 genes were expressed at least one time point during this developmental process. These genes were clustered based on expression profiles and during early stages of cyst formation, more genes were upregulated than downregulated compared to their expression profiles in trophozoites (time = 0 h). During later stages of encystations, the reverse was true and more genes were downregulated compared to trophozoites. This suggests a developmental program for cyst formation being activated during early stages of the cyst formation process. Based on which genes were differentially regulated, it was clear that main metabolic enzymes were being shut down in the later stages of cyst formation, probably in preparation for the dormant cyst stage.[31] This work was also compared to two other studies focusing on encystations in *Entamoeba*, a proteomic study[33] and a metabolomic study.[34] Both studies were somewhat limited in scope compared to the Ehrenkaufer et al. study and only some overlap was identified. This was partly due to the fact that proteomic study was conducted on *E. histolytica* and possible orthologs were difficult to identify. The major discovery was the possible important role of phospholipase D (PLD) in very early cyst-forming stages. PLD is involved in membrane events and cleaves the phosphodiester bond in structural phospholipids. This results in the production of phosphatidic acid

Downloaded From Vet-eBooks.com

that can act as a secondary messenger and can relay intracellular signals.[35] Functional experiments demonstrated indeed that cyst formation is affected when PLD is inhibited by adding n-butanol to the induction medium.[31] The work by Ehrenkaufer et al. has produced a working-list of possible essential encystation genes, and further work, both in *E. invadens* and *E. histolytica*, will be required to see if the life cycle of this important parasite can be disrupted.

4. Single-Cell Technologies

Since the previous edition of this book was published, massive progress has been made in the single-cell sequencing field. Up to now, most sequencing efforts focused on multiple cells. The yeast genome[9] was not the genome from one yeast cell but from a culture. The human genome[36,37] was not from one cell either. Obviously, the ability to sequence single cells stands or falls with the ability to reliably isolate one cell and then obtain enough material to do meaningful sequencing. Despite a quick start, a realization is now crystallizing that at many stages artifacts can be introduced and that some startling papers have perhaps been too hasty in celebrating their achievements.[38] One of the big challenges has been the isolation of single cells. Cells can be obtained by enzymatic digestion to remove them from surrounding tissue, by laser-capture microdissection, by fluorescence-activated cell sorting (FACS), or by microfluidics.[39] It is important that wells are inspected as to make sure there is a cell in a well or to be certain there is not more than one cell in each well to avoid erroneous results creeping in at this early stage. The next critical step that is prone to artifact introduction is the whole genome amplification step. Artifacts such as chimeric sequences, mutations, and amplification errors can seriously hamper identification of genuine sequence diversity. Principally, there are two methods, a PCR-based approach or an isothermal amplification approach (a third method is a combination of the two). For the PCR approach, the genome is amplified using known genome-wide sequences, an adaptor ligated to sheared DNA, or using random primers. The main problem with this approach is the nonrandom distribution of primer sites and/or PCR bias due to GC content.[39,40] The isothermal amplification method uses ψ29 polymerase which results in greater genome coverage with lower error rates. However, errors are introduced because initial amplification products get overrepresented in the final DNA.[39] In addition, some chimeric products are being produced as well. The hybrid method uses a combination of the PCR-based and isothermal approaches and produces intermediate results, and it is this method and the isothermal method that seemed to be most used.[39] However, the main problem remains that introduced artifacts will make it difficult to separate these from genuine genomic features. A pragmatic approach is to sequence at least three cells to discriminate artifacts from genuine base substitutions. Obviously, this will make the discovery of very rare variations quite challenging.[40]

An important rationale to conduct single-cell transcriptomics is that standard transcriptomics approaches (using populations of cells), will only result in trends gene

Downloaded From Vet-eBooks.com

expression. Mutually exclusive gene sets will never be discovered unless single-cell transcriptome data is available.[40] A key challenge in transcriptomics is that estimations suggest that only 5−25% of mRNAs are converted into cDNA.[41] In addition, it may be difficult to amplify or sequence extremely low-level transcripts, or, at the bioinformatics stage, the low number of reads representing these transcripts may be considered as noise by some software tools and undesirably removed from the dataset.[40] As direct sequencing approaches for nucleic acids are being developed and optimized,[42] perhaps the amplification steps that introduce artifacts are no longer needed in the (near) future.

Due to the small genome sizes, most viral genomes are readily sequenced (see the Ebola section in this chapter). Several studies benefitted from this and were able to follow genome evolution in populations and even within patients. The classic case of the Florida dentist and HIV transmission[43] and the more recent spread of Ebola[20] are all clear examples of live tracking of genome evolution. A 2014 study using direct sequencing of eukaryotes to understand variation within malaria patients[44] brings infectious disease research to a new level. People living in endemic areas of infectious diseases are often exposed to multiple infection events. Malaria patients bitten by a mosquito carrying different *Plasmodium* species or strains can develop complications as the new infection might carry drug resistance for example. Although these multiple-genotype infections are well known, their effect on drug resistance, virulence evolution, intrahost dynamics, and recombination rates is not known.[44] Nair et al. exploited the fact that *Plasmodium* lives inside erythrocytes to isolate single cells of the malaria parasite. Being aware of potential pitfalls, they tested 14 different combinations of cell sorting, whole genome amplification, and analytical methods to identify the most robust approach to perform single-cell genomics. This took 260 single-cell implications. Using rarefaction curves, the authors concluded that after 10−15 cells, they did not obtain new genotypes and suggested that sequencing more than 50 cells would not be recommended.[44] It needs reminding that this is dependent on the rate of change and number of infections present, and for other systems this number might be lower, or higher. Comparative analyses between single-cell genomes and deep sequence clonal genomes indicated that their approach effectively captured all present haplotypes. On average, for *Plasmodium vivax*, the authors obtained >30% coverage of the genome at a 10-fold sequence depth. For *P. falciparum*, this number was >50%. As they also deep sequenced several *P. falciparum* strains, they could match nearly 99% of all SNPs based on the single-cell genomes.[44] Interestingly, the original *P. falciparum* genome was affected by the parasite's high AT-content[13] but this sequence bias did not affect the single-cell genomes. As one of the questions about multiple-genotype infections is whether it affects spread of drug resistance, the authors compared alleles known to be involved in drug resistance within their single cell genomes. Although most alleles were fixed, some demonstrated intrahost polymorphism. One mutation was identified in a gene that has been linked to artemisinin resistance.[44] Although perhaps early days, but when direct nucleic acid sequencing becomes a routine method, real-time tracking and identification of infectious diseases within patients and personalized medicine will become a reality.

Downloaded From Vet-eBooks.com

5. High-Throughput Sequencing

Since mid-2000s, advances in sequencing technologies have made it possible for DNA sequencing to take place outside large dedicated sequencing centers (e.g., The Sanger Institute). Genomics studies are no longer restricted to the use of model species, and nonmodel environmental species can be relatively easily sequenced using a benchtop sequencer (e.g., Illumina MiSeq). It is now possible to sequence the entire genome of a bacterial or viral pathogen, assemble the raw sequence reads, perform automated annotation, and visualize the results within a couple of days to weeks. At the same time (indeed even on the same sequencer), it is also possible to selectively sequence the transcriptome (RNA-seq), regions of DNA bound to protein (ChIP-Seq), or, for relevant species, methylated DNA to study epigenetic effects, as well as small RNA molecules. It is also possible to perform the very same sequencing upon the host organism at the same time, facilitating studying host−pathogen interactions.

Bioinformatics algorithms and tools are crucial in analyzing such unprecedented volumes of data. These data volumes have emerged as a result of next-generation sequencers, such as the Roche/454, Illumina, and ABI/SOLiD systems. While useful information can be extracted by single researcher by targeted analysis of the sequencer output, to gain the most information out of such data, it is becoming increasingly common for multiple researchers or research groups with widely differing areas of expertise to collaborate. This collaboration is absolutely crucial if relevant insights are to be gained from large-scale datasets. As a result, a vast array of data is generated, which is required to be annotated and curated as well as analyzed for information relevant to any particular experiment. In addition, this information needs to be stored, shared, and distributed in a manner that enables reanalysis if and when new hypotheses are generated.

Experiments including high-throughput sequencing should be carefully designed and it is important that the bioinformatician(s) responsible for the data analysis are involved in this process to ensure, for example, that enough replicates are being included and/or that the genome or transcriptome to be sequenced is sufficiently covered and is at a suitable read depth. Not all pathogens can be cultured or isolated from host tissues and it is highly recommended to concentrate or (semi-)purify the pathogen before sequence library preparation. Even though it is possible to remove reads representing host sequences in silico, host DNA will be over-represented and sequenced, especially when the pathogen number in the tissue is low.

One of the first steps in the analysis of high-throughput sequences is the assessment of the quality of the reads. The most commonly used tool is FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), which processes the raw fastq sequence files and generates a report that shows various aspects of QC, including details on the number of reads, read length, Phred score quality, and GC-content along the whole read and potential adapter contamination. This provides a good overview of the overall quality of the dataset. In most cases, the reads will then need to be quality-trimmed, and many tools have been developed to achieve this (e.g., Trimmomatic[45] and Trim Galore!; http://www.bioinformatics.babraham.ac.uk/projects/trim_

Downloaded From Vet-eBooks.com

galore/). These software packages will remove any remaining adapter and low-quality sequences, trim leading and/or trailing sequences, and remove reads that are below a desired length. Many recent bioinformatics pipelines already include quality-trimmers and use error-correction methods, so it is not always required to do this separately.

6. *De Novo* Genome Assembly

A large array of programs is now available to assemble genomes and transcriptomes *de novo*. All of those are not listed in detail here as there are many considerations, including sequencing platform used, the read length in use, the expected genome size, length of longest repetitive elements, GC-content, and whether paired-end reads are in use. Examples of tools that have been optimized for bacterial and/or viral genome assemblies include the A5-MiSeq pipeline[46] for bacterial and viral genomes and the Iterative Virus Assembler (IVA[47]) , which assembles virus genomes (that have no repeat sequences) from mixed populations at extremely high and variable depth. When the genome contains regions of low complexity and/or repeat regions, significant gaps may be left in the final assembly when using short (paired) reads produced by, for example, Illumina sequencers only. For many analyses, especially for prokaryotic organisms, these gaps are generally not considered to be an issue as often the sequence information is used for the typing of bacterial strains based on a set of conserved genes (multilocus sequence typing), screening for the presence of resistance genes, and/or the identification of the core- and pan-genome. However, in cases where closure of these gaps is desirable, the addition of long sequence reads produced by, for example, PacBio or MinION sequencing can significantly improve the quality of the final genome assembly. The genomes of the emerging oomycete pathogen, *Pythium insidiosum*, and *Staphylococcus aureus* Tager 104 were sequenced and assembled using a combination of Illumina and PacBio reads and assembled using SPAdes.[48–50] Koren and Phillippy[51] reviewed the three leading long-read sequencing technologies (PacBio, Illumina synthetic long-read sequencing,previously known as Moleculo, and Oxford Nanopore MinION), their characteristics, and the algorithms that are currently available for long-read assembly.

7. Whole-Genome Sequence Analysis

Many different types of analysis can be performed using whole-genome sequence (WGS) data, including whole-genome annotation, strain typing using multilocus sequence typing (MLST) or WGS phylogenetic analysis, detection of variants (e.g., SNPs and indels), identification of the core-and pan-genome (as well as recombining regions in bacterial strains), and phylogeography analysis.

A selection of tools is available to annotate assembled genomes, including Prokka, a command-line software tool that uses a combination of gene and motif prediction tools to rapidly annotate prokaryotic and viral draft genomes.[52] It produces a set of standard output files, including FASTA files containing all identified gene and protein sequences,

Downloaded From Vet-eBooks.com

as well as Genbank and GFF files. Online annotation tools are also available, including RAST, which can be used to annotate bacterial and archaeal (draft) genomes.[53]

Since the late 1990s, MLST analysis has been used to characterize bacterial isolates[54] and with the ever-increasing number of bacterial sequences being generated, this type of analysis will become more specific. MLST analysis can be performed using BLAST searches against MLST databases derived from http://pubmlst.org/, either using custom scripts or as part of larger analysis pipelines, for example, nullarbor (https://github.com/tseemann/nullarbor).

Phylogenetic analysis based on whole bacterial or viral genome alignments can be performed to classify different species and, when used for bacterial sequence typing, it is even more accurate than MLST analysis as it uses all of the sequence information for comparisons, rather than a small selection of conserved loci. The first step is to identify the core-genome (the part of the genome that is shared across all strains/isolates), which can be determined using, for example, the Harvest suite,[55] followed by phylogenetic inference (using, e.g., RAxML[56]). Variants (SNPs and indels), potentially explaining virulence or resistance of particular pathogenic strains/isolates, can be identified by mapping reads to a reference genome sequence (e.g., by using BWA[57] or Bowtie),[58] viewed using genome visualization tools (e.g., IGV[59] or Tablet),[60] and filtered using variant calling software tools (e.g., SAMtools,[61] Varscan 2,[62] snippy (https://github.com/tseemann/snippy), or the GATK pipeline).[63]

Nullarbor (https://github.com/tseemann/nullarbor) is a bioinformatics pipeline that performs many of the analyses described earlier and it cleans Illumina sequence reads and performs *de novo* assembly of bacterial isolates. In addition, it attempts to identify the species (by comparison to all known bacterial genome sequences) and species type (using MLST analysis), variants (SNPs), and resistance genes, and the results are then summarized in a comprehensive report.

While studying bacterial genome sequences, further characterization analysis can be done, including the identification of the core- and pan-genomes across strains (e.g., using Roary)[64]; identification of recombining regions (e.g., clonalFrameML)[65]; and identification of antimicrobial-resistant genes (e.g., Ref. 66). Furthermore, comprehensive phylogeography analysis can be performed to study genome evolution of pathogens during outbreaks using BEAST.[67] Great examples of applications of the latter method are demonstrated in studies on the outbreak of *Legionella pneumophila* in Spain over an 11-year period[68] and the Ebola virus disease epidemic in West Africa.[19]

As was mentioned earlier, it is crucial that data are shared among the scientific community. An excellent example of an online tool that facilitates the sharing, visualization, and exploration of WGS data using trees, maps, and timelines is microreact (http://microreact.org/).

8. RNA-Seq (Transcriptomics)

Gene expression analysis (transcriptomics) can be used to study host−pathogen interactions and to identify important cellular processes or metabolic pathways that are

Downloaded From Vet-eBooks.com

important during the infection process. When a reference transcriptome is available for the species of interest, reads derived from different samples (e.g., infected vs. noninfected or infected tissue samples over time) can be mapped to this transcriptome and abundances for each gene compared in order to identify differentially expressed genes/transcripts. When a reference transcriptome is not available, a *de novo* transcriptome can be generated. *De novo* transcriptome assembly involves the reconstruction of gene transcripts, which is more complex compared to genome assembly because of the presence of isoforms and alternatively spliced transcripts. The Trinity RNA-Seq *de novo* assembly software package[69] is a widely-used pipeline for *de novo* assembly of transcriptomes and it is well supported by the developers. Trinity contains a selection of scripts that allow the user to generate a *de novo* transcriptome from raw sequence reads, estimate read abundance across a selection of samples, and perform differential gene/transcript expression analysis. It uses the most commonly used tools, including RSEM[70] and Kallisto[71] for read mapping and abundance estimation and edgeR,[72] limma,[73] and DESeq2[74] for differential gene/transcript expression analysis. The transcripts generated by Trinity can be annotated using Trinotate (https://trinotate.github.io/), which in turn contains a number of software tools to perform homology searches against public gene and protein sequences, screen for conserved protein motifs, and identify transmembrane regions and signal peptides.

Although progress in next-generation sequencing is fast, software development is even faster and the sell-by date of this section is likely to be in the near future. However, some software tools have been around for quite some time and so will several newer ones. Anyone interested in embarking on a next-generation project should not only look at the technical abilities of their sequencer of choice, but also include a bioinformatician from the start to make sure data can be analyzed.

9. Concluding Remarks

Remarkable progress has been made in the last decade due to enormous technological advancement in the sequencing field. Next-generation sequencers are now commonplace in many research institutes, and several commercial services provide relatively cheap sequence solutions. Our treatment of this field is not exhaustive and several good reviews about the technicalities of new sequencing platforms have been written on the topic.[75,76] We have attempted to show some recent case studies in infectious disease research that have been published in recent years using various next-generation sequencing approaches. In addition, the final sections provide a brief overview of bioinformatics solutions to the overwhelming data flow that next-generation sequencing platforms produce. Especially this last field is very fast-moving and new software solutions are written regularly. Many of the packages are community supported and those who are interested are invited to visit websites and Internet fora to familiarize themselves with this field. Despite the current pressures in science to produce and chase impact, the community spirit that once was synonymous with science is still present on the Internet. Embrace the technology that can help make a change in

Downloaded From Vet-eBooks.com

infectious disease research, but please remember to now and then look down a microscope to remember what parasites look like!

References

1. Hogeweg P. Simulating the growth of cellular forms. *Simulation* 1978;**31**:90−6.
2. Hogeweg P, Hesper B. Interactive instruction on population interactions. *Comput Biol Med* 1978;**8**:319−27.
3. Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J, et al. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature* 1976;**260**(5551):500−7.
4. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes JC, et al. Nucleotide sequence of bacteriophage phiX174 DNA. *Nature* 1977;**265**(5596):687−95.
5. Fiers W, Contreras R, Haegeman G, Rogiers R, Van de Voorde A, Van Heuverswyn H, et al. Complete nucleotide sequence of SV40 DNA. *Nature* 1978;**273**(5658):113−20.
6. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 1995;**269**(5223):496−512.
7. Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, et al. The minimal gene complement of *Mycoplasma genitalium*. *Science* 1995;**270**(5235):397−403.
8. Tomb J-F, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, et al. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 1997; **388**(6642):539−47.
9. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, et al. Life with 6000 genes. *Science* 1996;**274**(5287):546−67.
10. The-C.-elegans-Sequencing-Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 1998;**282**(5396):2012−8.
11. Organization WH. *World malaria report 2015*. Geneva (Switzerland): World Health Organization; 2015.
12. Liu W, Li Y, Learn GH, Rudicell RS, Robertson JD, Keele BF, et al. Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. *Nature* 2010;**467**(7314):420−5.
13. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 2002;**419**(6906):498−511.
14. Otto TD, Rayner JC, Böhme U, Pain A, Spottiswoode N, Sanders M, et al. Genome sequencing of chimpanzee malaria parasites reveals possible pathways of adaptation to human hosts. *Nat Commun* 2014;**5**.
15. Cheng Q, Cloonan N, Fischer K, Thompson J, Waine G, Lanzer M, et al. *Stevor* and *rif* are *Plasmodium falciparum* multicopy gene families which potentially encode variant antigens. *Mol Biochem Parasitol* 1998;**97**(1−2):161−76.
16. Gulia-Nuss M, Nuss AB, Meyer JM, Sonenshine DE, Roe RM, Waterhouse RM, et al. Genomic insights into the *Ixodes scapularis* tick vector of Lyme disease. *Nat Commun* 2016;**7**.
17. Corral-Rodríguez MÁ, Macedo-Ribeiro S, Barbosa Pereira PJ, Fuentes-Prior P. Tick-derived Kunitz-type inhibitors as antihemostatic factors. *Insect Biochem Mol Biol* 2009; **39**(9):579−95.
18. Korený L, Lukes J, Oborník M. Evolution of the haem synthetic pathway in kinetoplastid flagellates: an essential pathway that is not essential after all? *Int J Parasitol* 2010;**40**(2): 149−56.

Downloaded From Vet-eBooks.com

19. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* 2016;**530**(7589):228−32.
20. Gire SK, Goba A, Andersen KG, Sealfon RS, Park DJ, Kanneh L, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* 2014;**345**(6202):1369−72.
21. Westermann AJ, Forstner KU, Amman F, Barquist L, Chao Y, Schulte LN, et al. Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. *Nature* 2016; **529**(7587):496−501.
22. Hine PM, Bower SM, Meyer GR, Cochennec-Laureau N, Berthe FC. Ultrastructure of *Mikrocytos mackini*, the cause of Denman Island disease in oysters *Crassostrea* spp. and *Ostrea* spp. in British Columbia, Canada. *Dis Aquat Organ* 2001;**45**(3):215−27.
23. Carnegie RB, Meyer GR, Blackbourn J, Cochennec-Laureau N, Berthe FC, Bower SM. Molecular detection of the oyster parasite *Mikrocytos mackini*, and a preliminary phylogenetic analysis. *Dis Aquat Organ* 2003;**54**(3):219−27.
24. Burki F, Corradi N, Sierra R, Pawlowski J, Meyer GR, Abbott CL, et al. Phylogenomics of the intracellular parasite *Mikrocytos mackini* reveals evidence for a mitosome in Rhizaria. *Curr Biol* 2013;**23**(16):1541−7.
25. Adl SM, Simpson AG, Farmer MA, Andersen RA, Anderson OR, Barta JR, et al. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol* 2005;**52**(5):399−451.
26. van der Giezen M. Evolution: one thread to unite them all. *Curr Biol* 2013;**23**(16): R679−81.
27. Gaston D, Tsaousis AD, Roger AJ. Predicting proteomes of mitochondria and related organelles from genomic and expressed sequence tag data. *Methods Enzymol* 2009;**457**: 21−47.
28. Netz DJ, Mascarenhas J, Stehling O, Pierik AJ, Lill R. Maturation of cytosolic and nuclear iron-sulfur proteins. *Trends Cell Biol* 2014;**24**(5):303−12.
29. Ali IK, Clark CG, Petri Jr WA. Molecular epidemiology of amebiasis. *Infect Genet Evol* 2008;**8**(5):698−707.
30. Siegesmund MA, Hehl AB, van der Giezen M. Mitosomes in trophozoites and cysts of the reptilian parasite *Entamoeba invadens*. *Eukaryot Cell* 2011;**10**:1582−5.
31. Ehrenkaufer GM, Weedall GD, Williams D, Lorenzi HA, Caler E, Hall N, et al. The genome and transcriptome of the enteric parasite *Entamoeba invadens*, a model for encystation. *Genome Biol* 2013;**14**(7):R77.
32. Loftus B, Anderson I, Davies R, Alsmark UC, Samuelson J, Amedeo P, et al. The genome of the protist parasite *Entamoeba histolytica*. *Nature* 2005;**433**(7028):865−8.
33. Ali IKM, Haque R, Siddique A, Kabir M, Sherman NE, Gray SA, et al. Proteomic analysis of the cyst stage of *Entamoeba histolytica*. *PLoS Negl Trop Dis* 2012;**6**(5):e1643.
34. Jeelani G, Sato D, Husain A, Escueta-de Cadiz A, Sugimoto M, Soga T, et al. Metabolic profiling of the protozoan parasite *Entamoeba invadens* revealed activation of unpredicted pathway during encystation. *PLoS One* 2012;**7**(5):e37740.
35. Kolesnikov YS, Nokhrina KP, Kretynin SV, Volotovski ID, Martinec J, Romanov GA, et al. Molecular structure of phospholipase D and regulatory mechanisms of its activity in plant and animal cells. *Biochem Mosc* 2012;**77**(1):1−14.
36. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science* 2001;**291**(5507):1304−51.
37. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;**409**(6822):860−921.

Downloaded From Vet-eBooks.com

38. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 2014;**12**:87.
39. Gawad C, Koh W, Quake SR. Single-cell genome sequencing: current state of the science. *Nat Rev Genet* 2016;**17**(3):175−88.
40. Macaulay IC, Voet T. Single cell genomics: advances and future perspectives. *PLoS Genet* 2014;**10**(1):e1004126.
41. Islam S, Kjallquist U, Moliner A, Zajac P, Fan J-B, Lonnerberg P, et al. Highly multiplexed and strand-specific single-cell RNA 5′ end sequencing. *Nat Protoc* 2012;**7**(5):813−28.
42. Coupland P, Chandra T, Quail M, Reik W, Swerdlow H. Direct sequencing of small genomes on the Pacific Biosciences RS without library preparation. *Biotechniques* 2012; **53**(6):365−72.
43. Hillis DM, Huelsenbeck JP. Support for dental HIV transmission. *Nature* 1994;**369**(6475): 24−5.
44. Nair S, Nkhoma SC, Serre D, Zimmerman PA, Gorena K, Daniel BJ, et al. Single-cell genomics for dissection of complex malaria infections. *Genome Res* 2014.
45. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;**30**(15):2114−20.
46. Coil D, Jospin G, Darling AE. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* 2015;**31**(4):587−9.
47. Hunt M, Gall A, Ong SH, Brener J, Ferns B, Goulder P, et al. IVA: accurate *de novo* assembly of RNA virus genomes. *Bioinformatics* 2015;**31**(14):2374−6.
48. Ascunce MS, Huguet-Tapia JC, Braun EL, Ortiz-Urquiza A, Keyhani NO, Goss EM. Whole genome sequence of the emerging oomycete pathogen *Pythium insidiosum* strain CDC-B5653 isolated from an infected human in the USA. *Genom Data* 2016;**7**:60−1.
49. Davis RW, Brannen AD, Hossain MJ, Monsma S, Bock PE, Nahrendorf M, et al. Complete genome of *Staphylococcus aureus* Tager 104 provides evidence of its relation to modern systemic hospital-acquired strains. *BMC Genomics* 2016;**17**(1):179.
50. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;**19**(5):455−77.
51. Koren S, Phillippy AM. One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Curr Opin Microbiol* 2015;**23**:110−20.
52. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;**30**(14): 2068−9.
53. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 2008;**9**:75.
54. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* 1998;**95**(6):3140−5.
55. Treangen TJ, Ondov BD, Koren S, Phillippy AM. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol* 2014;**15**(11):524.
56. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;**30**(9):1312−3.
57. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;**25**(14):1754−60.
58. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009;**10**(3):R25.

Downloaded From Vet-eBooks.com

59. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 2013;**14**(2): 178−92.
60. Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, Cardle L, et al. Using Tablet for visual exploration of second-generation sequencing data. *Brief Bioinform* 2013;**14**(2):193−202.
61. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 2011; **27**(21):2987−93.
62 Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 2012;**22**(3):568−76.
63. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;**20**(9):1297−303.
64. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;**31**(22):3691−3.
65. Didelot X, Wilson DJ. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput Biol* 2015;**11**(2):e1004041.
66. Rowe W, Baker KS, Verner-Jeffreys D, Baker-Austin C, Ryan JJ, Maskell D, et al. Search engine for antimicrobial resistance: a cloud compatible pipeline and web interface for rapidly detecting antimicrobial resistance genes directly from sequence data. *PLoS One* 2015;**10**(7):e0133492.
67. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 2012;**29**(8):1969−73.
68. Sanchez-Buso L, Comas I, Jorques G, Gonzalez-Candelas F. Recombination drives genome evolution in outbreak-related *Legionella pneumophila* isolates. *Nat Genet* 2014;**46**(11): 1205−11.
69. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 2011;**29**(7):644−52.
70. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011;**12**:323.
71. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal RNA-Seq quantification. *arXiv* 2015 (1505.02710v2 [q-bio.QM]).
72. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;**26**(1):139−40.
73. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015; **43**(7):e47.
74. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**(12):550.
75. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. *Trends Genet* 2014;**30**(9):418−26.
76. Mardis ER. Next-generation sequencing platforms. *Annu Rev Anal Chem* 2013;**6**(1): 287−303.

Downloaded From Vet-eBooks.com

Genomics of Infectious Diseases and Private Industry

18

G. Vernet
Centre Pasteur, Yaoundé, Cameroon

1. Introduction

Genomics is the science of studying genomes of living organisms. The main driver of companies that specialize in genomics is human-genome sequencing and most of the technology breakthroughs have been achieved to improve the speed and quality of human whole-genome sequencing (WGS) and to reduce its cost and turnaround time. However, noneukaryotic genomics, and especially pathogen genomics, have made considerable progress since mid-1980s, and the first whole genome ever sequenced was from an RNA-virus, the bacteriophage MS2.[1] This was just 1 year before the invention of dideoxy chain terminator sequencing by Fred Sanger, which was the first breakthrough in pathogen genomics. The first bacterial whole-genome sequence was obtained in 1995 from *Hemophilus influenza*.[2] Applied Biosystems Inc. (ABI), now one of the various brands of Thermo Fisher Scientific Inc. (Middlesex, MA, USA), has been the most renowned company providing sequencing instruments and reagents during nearly 2 decades. This was the era of de novo reference quality genome sequences.

The availability of such reference genomic data for prokaryotic and eukaryotic organisms has led to the second breakthrough in pathogen genomics: genome resequencing using microarrays which allowed the comparison with a reference genome of specific species and the detection of variants: single nucleotide polymorphisms (SNPs), insertions, or deletions with a potential impact on the phenotype. As of February 2016, 31,633 scientific articles using the Affymetrix (Santa Clara, CA, USA) proprietary microarray technology have been listed on the website of the company (www.affymetrix.com). Affymetrix still proposes whole-genome genotyping chips, the Genome-Wide Human SNP Array 5.0 that provides information on 500,000 SNPs and also contains 420,000 probes to provide additional data such as copy number variation. However, microarray technologies are now mostly used for targeted resequencing of specific genes, most often in a custom-based approach.

The microarray technologies have been supplanted by next-generation sequencing (NGS). NimbleGen, now part of Roche Sequencing (Pleasanton, CA, USA), manufactured DNA chips from 1999 but stopped this activity in 2012, and Affymetrix has been recently acquired by Thermo Fisher Scientific. NGS represents the third breakthrough in pathogen genomics and has brought high throughput to sequencing. As it is the case for DNA chips, the short length of reads generated by the first NGS technologies, either sequencing-by-synthesis (SBS) or semiconductor sequencing, makes them

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00018-4
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

resequencing technologies rather than de novo sequencing technologies. However, NGS has been a real revolution. Sanger sequencing was used for the 13-year-long Human Genome Project, which resulted in the first whole-genome sequence in 2003 for a budget of $2.7 billion. Five years after, the same result was obtained in 5 months for just $1.5 million.[3] These progresses were largely due to the introduction of the first NGS platform by 454 Life Sciences (now part of Roche Sequencing) in 2005. Interestingly, NGS technologies and instruments were first used to sequence the whole genome of human pathogens, *Mycoplasma genitalia*[4] and *Escherichia coli*.[5] Because of the high throughput of NGS platforms, as of February 2016, more than 75,000 complete genome sequences from 5375 viruses and more than 2000 complete bacteria genomes from more than 500 genera are available in the NCBI database (http://www.ncbi.nlm.nih.gov/), and these numbers grow rapidly, the emphasis being on species that are pathogenic for animals or plants.

The National Human Genome Research Institute (NHGRI) part of the NIH awarded in 2008 more than $20 million in grants (the $1000 genome grants) to develop innovative sequencing technologies inexpensive and efficient enough to sequence a person's DNA as a routine part of biomedical research and health care for less than $1000. The cost of a human WGS dropped dramatically and, as of October 2015, was estimated by NHGRI at $1245 (Fig. 18.1). The same cost drop applies to smaller genomes such as those of microorganisms as the cost to sequence 1 Mb is now estimated at $0.014 and a target for the sequencing of a bacterial

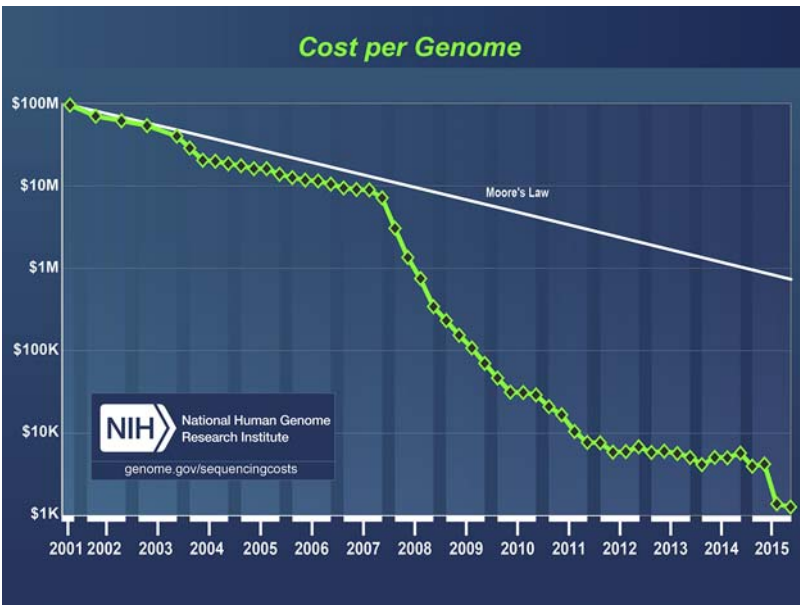


Figure 18.1 Human whole genome cost drop over the years.
Data from the National Human Genome Research Institute, NIH, USA, http://www.genome.gov/sequencingcosts/ (last visited February 11, 2016).

Downloaded From Vet-eBooks.com

genome could be as low as $1.[6] Currently, the cost of sequencing a bacterial genome is between $140 and $180.[7]

In 2011, the median turnaround time to obtain a human whole-genome sequence was just 68 days (http://www.completegenomics.com/). The turnaround time of bacterial WGS is now around 1 day and Ebola virus could recently be sequenced from a clinical specimen in less than 24 h, the sequencing itself being completed in 15−60 min.[8]

The fourth breakthrough in genomics is the single-molecule real-time sequencing (SMRT) technologies proposed by companies such as Pacific Biosciences (Menlo Park, CA, USA) and Oxford Nanopore (Oxford, UK; http://www.nanoporetech.com) since 2014, which achieve high-throughput long-read length sequencing. With these techniques, still less accurate than high-throughput, short-read length sequencing, genomics comes back to the era of whole-genome reference sequences.

All manufacturers now target benchtop format instruments, and small-sized and portable sequencers may allow sequencing in clinical laboratories for diagnosis as well as "in the fields," during large epidemics for genomic surveillance.

Of course, massively parallel sequencing generates enormous amounts of data and bioinformatics technologies for quality control, genome assembly, comparison with reference genomes, sequence mapping, and annotation, and analysis of sequence variations has also evolved together with sequencing technologies. "Big data" must, in addition, be stored somewhere and companies now propose cloud computing solutions.

Understanding hosts, vectors, and pathogens' genomes, as well as their transcriptomic and epigenomics modifications following infection, during the course of the disease, and under treatment will ultimately lead to personalized medicine for which genome characteristics will be important to tailor treatments.[9]

Several biotech companies have been created since 1990s to develop and market systems for the analysis of genomes, many of them by scientists issued from universities. Since then, we witness a very important concentration of technologies, both NGS and microarrays, in the hands of few large manufacturing or service companies that propose solutions for chemistry, instruments, and software for data analysis.

2. Technologies and Instrument Platforms

2.1 Sanger Sequencing

Sanger sequencing uses the SBS approach in which a DNA polymerase generates DNA reads from a template that is the DNA molecule to be analyzed. The nature of the nucleotide at a given position is now determined using specific dyes.

Sanger sequencing, although too laborious and expensive for WGS, remains routinely used when sequencing of specific genes or fragment of genes is needed, for example, for viral or bacterial genotyping or for resistance testing when SNPs

Downloaded From Vet-eBooks.com

are associated with specific genome regions. For bacterial WGS, biological amplification by culture and single colony picking is needed whereas PCR amplification of specific genes is done for both viruses and bacteria before amplicons are sequenced. Since 1987 and during the last four decades, Sanger sequencing has been mostly done on ABI sequencers (Thermo Fisher Scientific) instruments, a brand that now proposes a series of capillary electrophoresis sequencers ranging from 1 to 96 capillaries and covering the needs of different laboratories in terms of throughput. All current ABI DNA sequencing kits use cycle sequencing protocols with two different chemistries: dye primer chemistry or dye terminator chemistry.

2.2 Next-Generation Sequencing Based on Sequencing-by-Synthesis

NGS is parallel sequencing of clonally amplified or single DNA molecules by iterative cycles of polymerase-based extension or oligonucleotide ligation that takes place in flow cells. These technologies have revolutionized all aspects of genomics:

WGS, targeted resequencing, metagenomics, gene expression profiling, epigenomics, and DNA−protein interactions study (ChIP-Seq). The main characteristics of NGS technologies and platforms are described in Table 18.1. For a detailed description of NGS technologies, see Metzker M.L. (2010).[10]

2.2.1 Short-Read Sequencing: Reversible Termination Sequencing-by-Synthesis

This method is in principle similar to Sanger sequencing although the chain termination process is rendered reversible by using special fluorescently labeled terminator nucleotides. This technology allows high throughput and accuracy and is the predominant method used in pathogen genomics.

454 Life Sciences (Roche Sequencing) has been the first company to commercialize an NGS platform in 2005. The company now sells two instruments using the same chemistry: GS FLX, which is recommended for genomic analysis of complex organisms or samples with multiple bacterial genomes and GS Junior, which is recommended for genomic analysis of viruses, bacteria, and fungi. The FLX system can perform high-throughput sequencing of 100s to 1000s of samples and loci, whereas the Junior system is limited to targeted sequencing of 10s to 100s of samples and loci. However, the latter benchtop instrument is more suited to microbiology laboratories. Automation using the magnetic beads technology simplifies emulsion-PCR and allows library preparation of genomics samples in hours in a single tube, eliminating cloning, and colony picking. Currently, 671 articles using the 454 technology for microorganism genomics have been published (http://sequencing.roche.com/).

Illumina (San Diego, CA, USA) was created in 1998 to exploit rights on the BeadArray technology developed at Tufts University. Its first NGS, the Genetic Analyzer II, platform has been released in 2007. Currently, the company proposes a complete range of systems including a small benchtop format (MiniSeq), MiSeq, especially designed for targeted and small genome sequencing (including the first

Downloaded From Vet-eBooks.com

Table 18.1 Main Characteristics of Current NGS Platforms

	Ion S5 XL (Thermo Fisher)				Ion PGM (Thermo Fisher)				MiSeq (Illumina)				GS Junior Plus System		GS FLX+		MinIION	PacBio RS II (Pacific Biosciences)
	Ion 520	Ion 530	Ion 540		Ion 314	Ion 316	Ion 318		Reagent Kit V2		Reagent kit V3		(Roche)		(Roche)		(Oxford Nanopore)	1 to 4 cells
Read length	200 bp/400 bp				200 bp/400 bp				36 bp/2x25 bp/2x150 bp/2x250 bp				700 bp		600 bp/1000 bp		10 Kb (up to 300 Kbp)	> 20 Kbp
Output	**200 bp** 0.6–1 Gb	3–4 Gb	10–15 Gb		**200 bp** 30–50 Mb	300–600 Mb	600 Mb–1Gb		**36 bp** 540-610 Mb		**2 × 75 bp** 3.3–3.8 Gb		**700 bp** 70 Mb		**600 bp** 450 Mb		6 Gb (48 hours run)	500 Mb/1 Gb per cell
	400 bp 1.2–2 Gb	6–8 Gb	–		**400 bp** 60–100 Mb	600 Mb–1Gb	1.2–2 Gb		**2 x 25 bp** 750-850 Mb		**2 × 300 bp** 13.2-15 Gb				**1,000 bp** 700 Mb			
									2 x 150 bp 4.5-5.1 Gb									
									2 x 250 bp 7.5-8.5 Gb									
Reads (million)	3–5	15–20	60–80		0,40–0,55	2–3	4–5.5		24-30		44-50		0.1		1		0.6	0,055 per cell
Quality scores	no data				no data				**36 bp**	> 90% bases higher than Q30	**2 × 75 bp**	> 85% bases higher than Q30	Q20 read length of 700 bases (99% accuracy at 700 bases and higher for preceding bases)		**600 bp**	99.995% (x 15 coverage)	up to 96%	> 99,999 (QV50)
									2x25 bp	> 90% bases higher than Q30	**2 × 300 bp**	> 70% bases higher than Q30			**1,000 bp**	99.997% (x 15 coverage)		
									2x150 bp	> 80% bases higher than Q30								
									2x250 bp	> 75% bases higher than Q30								
Analysis time	**200 bp** 1 hr	2.5 hr	5 hr		**200 bp** 2.3 hr	3.0 hr	4.4 hr		**36 bp** 4 hrs		**2 × 75 bp** 21 hrs		**700 bp** 18 hours		**600 bp** 10 hours		Real time	Real time
	400 bp 2 hr	4 hr	–		**400 bp** 3.7 hr	4.9 hr	7.3 hr		**2 x 25 bp** 5.5 hrs		**2 × 300 bp** 56 hrs				**1,000 bp** 23 hours		2 minutes to 48 hours	30 minutes to 6 hours
									2 x 150 bp 24 hrs									
									2 x 250 bp 39 hrs									
Dimension (W x D x H) cm	30.8 x 69.8 x 44.4				61 x 51 x 53				68,6 x 52,3 x 56,5				40 x 60 x 40		74.3 × 69.8 × 36.1		10,5 x 2,3 x 3,3	61.3 x 91.3 x 66.5

Bold means read length.
Data from manufacturers' websites.

Downloaded From Vet-eBooks.com

FDA-approved NGS instrument and a fully validated system for forensic genomics), the NextSeq and HiSeq series for higher throughput analysis, and WGS of complex organisms. After library preparation, cluster generation is done on the NeoPrep Library Prep System, which significantly reduces hands-on time compared to emulsion PCR. As of 2016, nearly 2500 articles have been published using Illumina technology for microbial genomics (www.illumina.com/).

Semiconductor sequencing is an alternative to nonoptical methods, such as 454 or Illumina, which reduces costs. The most popular system based on semiconductor is Ion Torrent with the Ion PGM System and Ion Proton systems (Thermo Fisher Scientific), which monitors the release of hydrogen ions, another by-product of DNA synthesis when a particular nucleotide is incorporated in the DNA chain. The Ion PGM system is recommended for de novo microbial sequencing, bacterial typing research, multilocus typing (MLST), and viral typing research.

More recently, the company has released a new system: The Ion S5 System that is dedicated to infectious diseases. This system enables sequencing of whole bacterial genomes from isolates or the direct sequencing of specific viral, bacterial, or fungal genes (e.g., 16S genes or antibiotic resistance genes) from biological samples without the need for culturing, in as little as 24 h.

For a detailed description of Ion Torrent semiconductor technology, see http://www.ebi.ac.uk/training/online/course/ebi-next-generation-sequencing-practical-course/what-next-generation-dna-sequencing/ion-torre.

2.2.2 *Single-Molecule Sequencing: Real-Time, Long-Read Sequencing-by-Synthesis*

Single-molecule, long-read sequencing is the newest approach of NGS. The first commercially available system, the PacBio RSII system, was released by Pacific Biosciences (Menlo Park, CA, USA; http://www.pacb.com/) in 2011. In this system, a single DNA molecule is added to a well with a single DNA polymerase molecule. Nucleotides with a dye molecule attached to its phosphate are used and allow uninterrupted DNA synthesis. This real-time approach generates reads that are hundreds of times longer than those obtained with other NGS techniques. Even if throughput is lower, single-molecule, real-time sequencing has been used in applications where accurate long sequence reads are required, such as the project of Public Health England in partnership with the Wellcome Trust Sanger Institute to sequence its complete collection of bacterial strains (http://www.sanger.ac.uk/resources/downloads/bacteria/nctc/).

2.3 *Nanopore-Based Sequencing*

This technology which also generates single-molecule long-reads does not rely on SBS but on the use of nanopores through which a single DNA molecule passes. Each base going through the nanopore deviates an electrical current to an extent that is specific and measurable so that the technique can determine the sequence. Oxford Nanopore Technologies (Oxford, UK) has recently released the MinION, a USB drive−sized

Downloaded From Vet-eBooks.com

platform, which is based on that principle. It has been tested in many laboratories through an access program[11] and has recently been used to sequence Ebola virus from 142 samples collected during the recent epidemics. Results were generated in less than 24 h (15−60 min only for sequencing). This short turnaround time combined with portability of the equipment makes MinION particularly suitable for real-time genomic surveillance and outbreak monitoring.

The main characteristics and performances of NGS platforms are detailed in Table 18.1. The data presented here are provided by manufacturers on their websites. However, performances claimed by manufacturers may be overestimated. Harismendy et al.[12] have compared the platforms of 454 Life Sciences, Illumina, and Sanger Applied Biosystems on a 260 kb human-genome sample. Using the current versions of these platforms, although the Illumina and Applied Biosystems produced the largest amounts of data, only 43% and 34% of them, respectively, are usable after quality filtration. In contrast, 95% of the data generated by the 454 platform were usable. As expected, ABI Sanger sequencing had an error rate of about 7%. The overall sequencing accuracy of NGS platforms was very high (99.99%), but the ability to detect variant was 95% for the 454 platform (which had the lowest sensitivity), 100% for the Illumina platform, and 96% for the ABI platform, the last two technologies being less specific.

2.3.1 Microarrays

At least 36 companies providing microarrays were identified in 2009. Most of them proposed low- to medium-density custom arrays, which are glass plates or beads with DNA probes either spotted or synthesized in situ using a variety of technologies, including printing with fine-pointed pins onto glass slides, photolithography using premade masks, photolithography using dynamic micromirror devices, inkjet printing, or electrochemistry on microelectrode arrays. For a review on the use of microarrays for clinical microbiology, see Miller and Tang.[13] Two companies were proposing high-density DNA-Chips, Affymetrix (Thermo Fisher Scientific), and NimbleGen (Roche Sequencing). Affymetrix still commercializes microarrays, but none is currently intended for the analysis of pathogen genomes whereas, in 2011, more than 1500 publications could be retrieved from the Affymetrix database with the key words “virus,” “bacteria,” “parasite,” and “fungi.” Roche NimbleGen stopped manufacturing microarrays in 2012 and is now focusing on SeqCap Target Enrichment brand for targeted DNA sequencing, bisulfite sequencing, and transcriptome sequencing.

2.4 Software

The different sequencing approaches required different bioinformatics treatments. The first wave of genomic data targeting de novo WGS required tools for genome assembly, gene calling, and annotation, such as Phred, Phrap, Glimmer, or Artemis.[14−16] NGS now rather compares an incomplete sequence to a reference sequence, and software such as Newbler, SOAPdenovo, or Velvet are designed to analyze short reads.[17]

Downloaded From Vet-eBooks.com

3. Customers and Their Needs

Customers of companies involved in pathogen genomics are both public and private organizations. Final users may be either research laboratories, clinical laboratories, forensic laboratories, public health laboratories, hospital laboratories involved in infection control, and pharmaceutical companies. However, because pathogen genomics integrates different competencies for preanalytical, analytical, and data analysis, and because investment needs in equipment and software are high, major customers of companies that commercialize genomics solutions are very often organizations—public or private—that provide services to the previously cited laboratories, such as the Institute for Genomic Research, the US Department of Energy's Joint Genome Institute, the Whitehead Institute and the Broad Institute in the USA, the Welcome Trust Sanger Institute in the UK, the Genoscope in France, or the Beijing Genomics Institute in China. However, because of the evolution toward benchtop formats, smaller laboratories tend to equip themselves and less and less outsource sequencing.

3.1 Research

The first need of scientists is to sequence microbe genomes and build databases of sequences. WGS has been made possible by using Sanger sequencing and now single-molecule sequencing. Comparative genome sequencing, and targeted resequencing allow the comparison of the sequence of whole genomes or specific genes to a reference sequence and to identify SNPs, insertions, and deletions (indels) as well as large chromosomal rearrangements (structure or copy number changes). Some of these polymorphisms determine phenotypic characteristics of a strain or an isolate.[18] Scientists will use these data for fundamental research: to annotate all genes, understand genome organization, classify species, and study their evolution. The knowledge of genome organization, combined with functional genomics, is the basis to understand pathological properties of infectious agents. Among downstream applications are the development of new drugs, vaccines, and diagnostics including the establishment of resistance and escape profiles.

A new field of investigation is the study of microbiomes or metagenomics,[19] which consists in the systematic sequencing of all nucleic acids in a given ecological "niche"—gut, upper respiratory tracts, and skin—to identify all microbes. This allows to characterize the "normal" flora, for example, at different ages in life, and interactions between this flora in pathogenic agents during infection, in particular the exchange of genetic material conferring resistance and virulence. Applications by pharmaceutical, diagnostic, and agrofood companies are very important.

Transcriptome analysis is the characterization of all coding and noncoding transcriptional activity in any organism without a priori assumptions through annotation of SNPs and mapping to reference genomes, characterization of transcript isoforms, regulatory RNAs, or splice junctions and determination of the relative abundance of transcripts (gene expression analysis). Analysis of differential gene expression is important in hosts and pathogens as well as in vectors of transmissible diseases (mostly

Downloaded From Vet-eBooks.com

insects). Human, plant, or animal cells can be studied when they are confronted to infection to identify the mechanisms targeted by the pathogen and those by which they resist infections. Pathogens' gene expression during infection of the cell is an important area of investigation as it may reveal pathological mechanisms and targets for new drugs.

Epigenome analysis is the study of chromatin structure and gene regulation by CpG methylation, histone modifications, or DNA−protein interactions.

3.2 Clinical Biology and Public Health

Sequencing is generally not the first-line technology to identify a bacterial infection except mycobacterial infections whereas viral diagnosis makes often use of molecular biology and sequencing. WGS will probably not add more to existing diagnosis techniques. On the contrary, metagenomics may prove useful in analyzing all sequences from a given sample and identifying pathogens that could be at the origin of the disease without any culture step. There have already been examples of the use of metagenomics for the identification of pathogens in severe gastrointestinal brain and lung diseases.[20−23] This approach may also identify new or unknown pathogens in clinical specimens. However, there are still a number of obstacles. First, the number of small sequences generated requires complex computational approaches that are costly and time consuming. It is also difficult to separate microbial DNA from human DNA and the amount of DNA from a given pathogen may be very low, requiring complex extraction methods to allow the high-depth coverage of its genome required for assembly. Finally, the detection of a genome is not sufficient to identify it as the true or unique cause of a disease.

The use of WGS may be of value for antimicrobial susceptibility testing. It was shown in a 2013 study that WGS of *Staphylococcus aureus* isolates was able to predict resistance to different antibiotics with a correlation of 97% with phenotypic testing.[24] Of course, this use of WGS requires prior culture and updated databases of resistance genes.

Genomics already finds applications in public health, for example, in the surveillance and control of emerging infectious control, including resistant pathogens. WGS sequencing of thousands of *Streptococcus pneumoniae* strains, which is done at the Sanger Institute can analyze and predict the impact of pneumococcal vaccination. Genomic epidemiology has already been proven useful, and probably easier to use than current typing methods, for outbreak investigations in the hospital or the community (see examples in Ref. 7). By giving hospitals an improved understanding of the genetic markers of virulence and resistance, this service can help them understand how bacteria are transmitted, while enabling better containment of an epidemic, limiting the spread of infectious agents and improving surveillance approaches.

3.3 Other Applications of Genomics

Genomics can be used by pharmaceutical companies for R&D and quality control. Of course, results generated by research laboratories using genomic technologies can find

Downloaded From Vet-eBooks.com

industrial applications such as new and better tailored drugs and vaccines. Reverse vaccinology is a good example. Pharmaceutical companies that need to verify sterility of injectable drugs and vaccines, for example, vaccines produced by viral culture, may take advantage of metagenomics analysis of their bulk solutions. Diagnostic companies can base the identification of new biomarkers for diagnosis or treatment monitoring disease on genomics.

The study of microbiomes also has a potential application for forensics: sequencing the "bacteriome" in traces left on can be used to identify people as the flora present on the skin is a signature depending on food habits, environment, and diseases.

4. Industry Landscape

Since the invention of Sanger sequencing in 1977, many start-up biotechnology companies have been created to develop and manufacture chemistry solutions and instrument platforms to automate sequencing. Most of them have been created by scientists or engineers from universities. As seen in Fig. 18.2, which presents a chronology of companies' creation along with release of sequencing platforms breakthrough of genomics in microbiology, many of these start-up have been incorporated in larger companies. Applied Biosystems Inc. has been acquired by Perkin Elmer and later, through its merger with Invitrogen, became Life Technologies, which in turn was acquired by Thermo Fisher Scientific. Thermo Fisher Scientific also acquired Ion Torrent and Affymetrix to become a major player in genomics. Similarly, Roche acquired 454 Life Sciences as well as NimbleGen. The third major genomics player is Illumina that claims on its website that its SBS chemistry is the most widely adopted NGS technology, generating approximately 90% of global sequencing data (http://www.illumina.com/). All these companies integrate bioinformatics and cloud computing in their offer. Several companies such as Helicos Biosciences Corporation (Madison, MA, USA) which was the first to release a single-molecule sequencing platform have disappeared. Some technologies such as the SOLiD NGS platform developed by Applied Biosystems or NimbleGen microarrays also no longer exist. However, new start-up companies emerged in the recent years that propose promising technologies, such as Oxford Nanopore or Pacific Biosciences, which already commercialize instruments, and others, such as NABSys (Providence, RI, USA; http://nabsys.com/) or BASE4 Innovation (Cambridge, UK; http://www.base4.co.uk/), currently develop new concepts for genomics. The current efforts of the whole genomics industry are obviously on performances in terms of read length, read depth and genome coverage, accuracy, quality of base call, throughput, and turnaround time. The trend is also to develop smaller, portable low-cost machines that will be especially suitable for applications in microbiology.

Research laboratories from universities or institutes have different needs, which were, up to now, impossible to address with a single technology or instrument. Due

Downloaded From Vet-eBooks.com

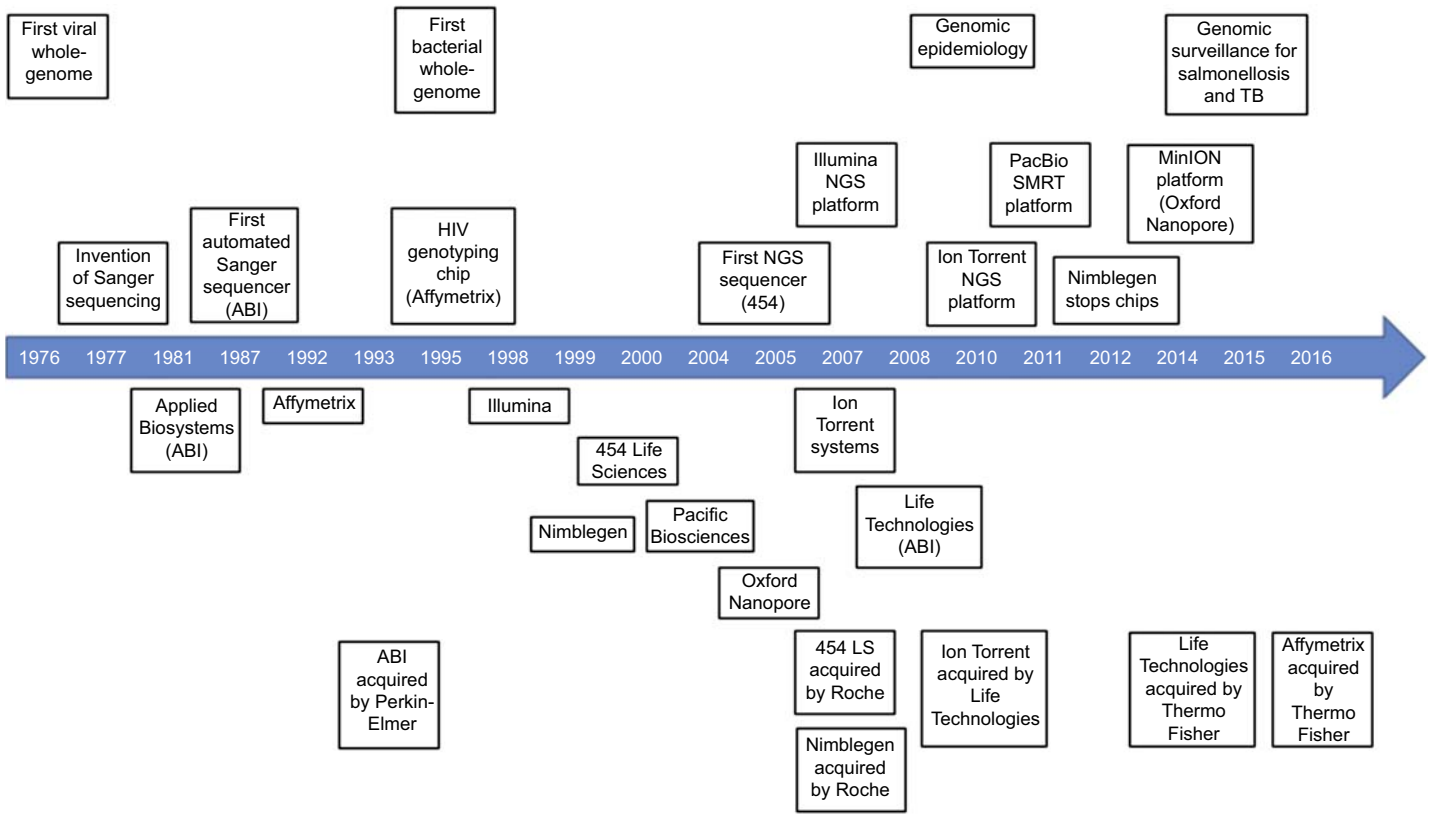


Figure 18.2 Chronology of genomics companies creation, NGS platform release, and major achievements of NGS in microbiology.

Downloaded From Vet-eBooks.com

to high prices of equipment and maintenance and frequent needs for instrument upgrade, platforms serving the needs of several research laboratories and providing services to the scientific community have emerged, which provide a full set of equipment and dedicated human resources. The Wellcome Trust Sanger Institute (Cambridge, UK) is an example where such a platform, linked to bioinformatics resources serves different research projects, including a pathogen genetic group exploring parasites, especially malaria and virus genomics. The Broad Institute (Cambridge, MA, USA) is another example with a genome-sequencing platform that also considers fungi, bacteria, and virus genomes. The J. Craig Venter Institute, a not-for-profit private organization based in Rockville, Maryland (USA), has more than 500 scientists and staff, more than 250,000 square feet of laboratory, and operates several resource centers (sequencing, genotyping, functional genomics, and bioinformatics) for infectious disease genomics. Probably the largest NGS service company, the Beijing Genomics Institute (BGI; Shenzhen, China), a spin-off from the Chinese Academy of Sciences, has more than 230 sequencing platforms, mostly from Illumina and Thermo Fisher Scientific (Ion Torrent) and has acquired, in 2013, Complete Genomics Inc. (Mountain View, CA, USA), a major supplier of DNA sequencing technology in the USA. BGI is a public−private company (http://bgi-international.com/). Such large institutes are not very numerous worldwide. However, they attract large international budgets that generate important sells for genomics companies in terms of instruments, maintenance, and reagents.

A nonexhaustive list of 228 companies and public facilities that provide DNA-sequencing services in different parts of the world can be found at http://grouthbio.com/Genome_Software_Service.php (last update in January 2016). Many of them are compliant with good clinical practices (GCP), good laboratory practices (GLP), and good manufacturing practices (GMP), or for clinical diagnostic services to the Clinical Laboratory Improvement Amendments (CLIA) regulations. Some of these companies have specialized in genomics of infectious diseases and provide expertise and R&D solutions to customers with particular needs, such as the identification of pathogens in clinical specimens with very low concentrations or in very degraded specimens (see, e.g., http://www.viroscan3d.com/).

bioMérieux (Marcy-l'Etoile, France), a world leader in the field of in vitro diagnostics, and Illumina recently announced the launch of bioMérieux EpiSeq, a service dedicated to the epidemiological monitoring and control of healthcare-associated infections. Hospitals facing a suspected epidemic or health crisis will be able to send bacterial isolates to a service laboratory designated by bioMérieux and equipped with an Illumina MiSeq sequencer. The genomic data is stored in a secure cloud platform and analyzed using the database and software developed by bioMérieux. Results showing the genomic profile of the infectious agents and the genetic variations identified by sequencing will be sent by bioMérieux to healthcare professionals in a customized, easy-to-interpret report. It is likely that in vitro diagnostic companies involved in infectious diseases will progressively enter the field of genomics with either manufacturing or services offers.

The total market of NGS was evaluated to be $484 million in 2008 (Vernet[25]). It was estimated at $3.3 billions in 2015 with a predicted annual growth rate of 21.3%

Downloaded From Vet-eBooks.com

between 2015 and 2020 (http://www.researchandmarkets.com/reports/3388723/global-next-generation-sequencing-ngs-market#pos-0).

5. Conclusion

Genomics has now found its place in all domains of activities in the field of infectious diseases, from basic to translational research, through disease diagnosis and surveillance, molecular epidemiology for outbreak investigation, and emerging infections monitoring. The current trend to instrument size reduction, portability, and cost and turnaround time reduction will change the landscape in the coming years by allowing smaller and decentralized laboratories to access NGS technology. Companies providing services will specialize on aspects that require strong R&D, such as specimen preparation and pathogen DNA enrichment or data analysis. In vitro diagnostic companies have started to enter the field and will probably take a growing part of the market. Manufacturing industries are experiencing concentration with a few leading companies, although innovation is still carried on by small biotech companies, often spin-offs of university laboratories. The market has a prediction of strong growth in the coming years.

References

1. Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J, et al. Complete nucleotide-sequence of bacteriophage MS2-RNA − primary and secondary structure of replicase gene. *Nature* 1976;**260**:500−7.
2. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 1995;**269**:496−512.
3. Voelkerding KV, Dames SA, Durtschi JD. Next-generation sequencing: from basic research to diagnostics. *Clin Chem* 2009;**55**:641−58.
4. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005;**437**:376−80.
5. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, et al. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 2005;**309**:1728−32.
6. Loman NJ, Pallen MJ. Twenty years of bacterial genome sequencing. *Nat Rev Microbiol* 2015;**13**:787−94.
7. Luheshi L, Raza S, Moorthie S, Hall A, Blackburn L, Rands C, et al. *Pathogen genomics into practice.* PGH Foundation, ISBN 978-1-907198-18-2.
8. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* 2016;**530**:228−32.
9. Snyder M, Du J, Gerstein M. Personal genome sequencing: current approaches and challenges. *Genes Dev* 2010;**24**:423−31.
10. Metzker ML. Sequencing technologies—the next generation. *Nat Rev Genet* 2010;**11**:31−46.

Downloaded From Vet-eBooks.com

11. Pennisi E. Genomics. Pocket DNA sequencers make real-time diagnostics a reality. *Science* 2016;**351**:800−1.
12. Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, et al. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol* 2009;**10**(3):R32.
13. Miller MB, Tang YW. Basic concepts of microarrays and potential applications in clinical microbiology. *Clin Microbiol Rev* 2009;**22**:611−33.
14. Rieder MJ, Taylor SL, Tobe VO, Nickerson DA. Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome. *Nucleic Acids Res* 1998;**26**:967−73.
15. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* 1999;**27**:4636−41.
16. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. *Bioinformatics* 2000;**16**:944−5.
17. Miller JR, Koren S, Sutton G. Assembly algorithms for next-generation sequencing data. *Genomics* 2010;**95**:315−27.
18. Herring CD, Palsson BØ. An evaluation of Comparative Genome Sequencing (CGS) by comparing two previously-sequenced bacterial genomes. *BMC Genomics* 2007;**8**:274.
19. Wooley JC, Godzik A, Friedberg I, Miller MB, Tang YW. Basic concepts of microarrays and potential applications in clinical microbiology. *PLoS Comput Biol* 2010:26−7.
20. Loman NJ, Constantinidou C, Christner M, Rohde H, Chan JZ, Quick J, et al. A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic *Escherichia coli* O104:H4. *JAMA* 2013;**309**:1502−10.
21. Rohde H, Qin J, Cui Y, Li D, Loman NJ, Hentschke M, et al. Open-source genomic analysis of Shiga-toxin-producing *E. coli* O104:H4. *N Engl J Med* 2011;**25**(365):718−24.
22. Wilson MR, Naccache SN, Samayoa E, Biagtan M, Bashir H, Yu G, et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N Engl J Med* 2014;**370**: 2408−17.
23. Fischer N, Rohde H, Indenbirken D, Günther T, Reumann K, Lütgehetmann M, et al. Rapid metagenomic diagnostics for suspected outbreak of severe pneumonia. *Emerg Infect Dis* 2014;**20**:1072−5.
24. Stoesser N, Batty EM, Eyre DW, Morgan M, Wyllie DH, Del Ojo Elias C, et al. Predicting antimicrobial susceptibilities for *Escherichia coli* and *Klebsiella pneumoniae* isolates using whole genomic sequence data. *J Antimicrob Chemother* 2013;**68**:2234−44.
25. Vernet G. *Genetics and evolution of infectious diseases*. 1st ed. London: Tibayrenc; 2011.

Downloaded From Vet-eBooks.com

Current Progress in the Pharmacogenetics of Infectious Disease Therapy

19

E. Elliot[1], T. Mahungu[2], A. Owen[1]
[1]University of Liverpool, Liverpool, United Kingdom; [2]Royal Free London NHS Foundation Trust, London, United Kingdom

1. Introduction

Following ingestion of standard doses of medication, interindividual variation in both desired and toxic effects is often observed. Factors contributing include age, gender, ethnicity, body mass index, physiological status, comorbidity, dietary factors, and coprescribed medication. The genetic contribution to interindividual variability was reported to range between 20% and 95%.[1] In infectious diseases, the most commonly studied variants are SNPs in genes implicated in drug absorption, distribution, metabolism, and excretion (ADME) pathways. There is increasing interest in nuclear receptors that regulate the expression of ADME genes, in human leukocyte antigen (HLA) subtypes in hypersensitivity reactions (HSRs), and in genes implicated in the development of metabolic toxicity.

2. Pharmacogenetics of HIV Therapy

The genetic contribution to variability in plasma concentration of antiretroviral drugs was assessed in 2015 using the relative genetic contribution (rGC) method.[2] Using this approach, the rank order for genetic contribution was efavirenz (EFV) > nevirapine (NVP) > etravirine > tenofovir disoproxil fumarate (TDF) > atazanavir (ATV) > ATV/ritonavir (ATV/r) > maraviroc (MVC) > lopinavir/ritonavir (LPV/r) > raltegravir (RAL), indicating that class-specific differences in rGC may exist.

2.1 Nucleoside/Nucleotide Reverse Transcriptase Inhibitors

Worldwide, there has been a move away from first-generation nucleosides, such as stavudine (d4T), zalcitabine (ddC), didanosine (ddI), and zidovudine (AZT), toward better-tolerated, less-toxic NRTIs, such as TDF, abacavir (ABC), and emtricitabine (FTC). First-generation nucleoside analogues are associated with the potential development of peripheral lipoatrophy, nonalcoholic steato-hepatitis, lactic acidosis,

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00019-6
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

pancreatitis, and peripheral neuropathy, through mitochondrial toxicity.[3,4] The newer generation NRTIs are better tolerated, but ABC is associated with HSRs, as well as cardiovascular disease (CVD) in some observational studies,[5,6] and TDF is associated with renal tubular toxicity[7] and bone density changes.[8] A newer tenofovir prodrug, tenofovir alafenamide (TAF), can be given at much lower doses and is associated with smaller declines in bone mineral density and more favorable changes in kidney biomarkers than TDF.[9]

ABC and zidovudine[10] are predominantly metabolized via hepatic glucuronidation while TDF and 3 TC undergo renal excretion. NRTIs are substrates for multidrug resistance protein 4 (MRP4/ABCC4), multidrug-resistant protein 5 (MRP5/ABCC5), and breast cancer-related protein (BCRP/ABCG2).[11−13] *ABCC4* variants have been reported to be associated with higher intracellular zidovudine triphosphate (*ABCC4* 3724G>A) and lamivudine triphosphate (*ABCC4* 4131T>G).[14]

HLA B*5701 is a strong predictor of the ABC HSR.[15] Caucasian patients receiving ABC have an 8% chance of developing potentially fatal HSRs within 6 weeks of initiating treatment.[7] The prospective screening for HLA B*5701 in a predominantly Caucasian population before commencing ABC has a high negative predictive value and significantly reduces the incidence of ABC HSR.[16] Although the carriage of HLA B*5701 and the subsequent rate of ABC HSR is less frequent in Black populations,[17,18] HLA B*5701 carriage has been reported to be 100% sensitive as a marker of immunologically confirmed ABC HSR in Black patients in the United States.[19] HLA B*5701 is, however, virtually absent among sub-Saharan Africans, and genetic testing is not recommended.[20]

The excretion of tenofovir in the kidneys is facilitated by both influx and efflux transporters, that is, human renal organic anion transporters (hOAT; SLC22 A),[21] located on the basolateral border of the proximal tubule and ABCC (MRP) transporters located on the brush border of the proximal renal tubule,[21,22] and the role of transporters for this drug was reviewed in 2014.[23] The *ABCC2* CATC haplotype (at positions −24, 1249, 3563, and 3972) and the allele −24C>T have both been associated with an increased incidence of tenofovir-associated tubular toxicity in predominantly Caucasian populations.[24] In a study in Japanese patients, the ABCC2 24T>C and 1249G>A were found to be protective for tenofovir-induced kidney toxicity,[25] but no association with glomerular filtration rate was observed in a Thai patient group.[26] The associations with *ABCC2* variants are difficult to rationalize, as tenofovir is not a substrate for ABCC2, although it may be that an endogenous substrate for ABCC2 exacerbates the toxicity of tenofovir or competes with tenofovir for transport by ABCC4 out of the cell. These genetic variants may also be in linkage disequilibrium with other SNPs in genes coding for unidentified factors exacerbating tenofovir toxicity.[23] Several studies have failed to identify an association between *ABCC4* polymorphisms and tenofovir renal toxicity.[24] *ABCC10* (coding for MRP7) polymorphisms have been associated with renal toxicity[27] and urinary to plasma tenofovir ratio.[28] Another study in Japanese patients did not replicate the association with renal toxicity,[25] but a case report of two vertically HIV-infected patients affected by kidney tubular dysfunction showed they both carried allelic variants in the *ABCC2*, *ABCC4*, or *ABCC10* genes.[29]

Downloaded From Vet-eBooks.com

2.2 Nonnucleoside Reverse Transcriptase Inhibitors

Until late 2000s, EFV was the preferred third agent in first-line regimens across resource-rich settings.[30,31] It has now been downgraded from recommended to alternative agent in favor of the protease inhibitors (PIs) and integrase inhibitors (InSTIs) in most guidelines.[32,33] It remains recommended as first-line third agent in resource-poor settings alongside NVP, which is also prescribed in the prevention of mother to child transmission (PMTCT).[34] Unfortunately, these drugs have a fragile genetic barrier to the development of resistance and a single-drug resistance mutation can confer high-level resistance to all first-generation NNRTIs.[35] The main difference between the two drugs lies in their toxicity profiles. The use of NVP is characterized by an idiosyncratic, potentially fatal, immune-mediated HSR, which occurs in about 5% of treated individuals during the first 6 weeks of therapy. This HSR can manifest as hepatotoxicity, fever and/or a Stevens–Johnson rash, and occasionally death. It is more likely in females with CD4 counts greater than 250 cells/mm^3 and in males with CD4 counts greater than 400 cells/mm^3.[36,37] The use of EFV on the other hand is characterized by the development of central nervous system (CNS) side effects in about 40% of treated patients.[38] Most of these symptoms (insomnia, dizziness, headache, and vivid dreams) are self-limiting and often disappear within the first 12 weeks of therapy,[39] but they can lead to treatment discontinuation and may also negatively impact upon compliance to therapy. However, the Encore I study in 2014 assessed the reduction of dose from 600 to 400 mg once daily and showed a lower incidence of adverse events without compromising efficacy.[40]

NVP is primarily metabolized by the cytochrome P450 3A4 (CYP3A4) and 2B6 (CYP2B6) enzymes into its major metabolites 2-hydroxyNVP and 3-hydroxyNVP, respectively, with a minor contribution from CYP3A5.[41] CYP3A4 (in concert with CYP2D6 and CYP3A5) is also involved in the biotransformation of NVP to 8-hydroxyNVP and 12-hydroxyNVP, both minor metabolites.[41] EFV on the other hand is predominantly metabolized by CYP2B6 into 8-hydroxyEFV with a minor contribution from CYP3A4.[42] The biotransformation of EFV to 7-hydroxyEFV, its minor metabolite, is via CYP2A6 with a minor contribution from CYP2B6.[42] Additionally, UDP-glucuronosyltransferase (UGT) 2B7 has been identified as the main enzyme involved in EFV N-glucuronidation and has been shown to predict EFV plasma concentrations.[43] The role of influx and efflux transporters in the disposition of NNRTIs is not fully characterized. Isolated studies have shown that both drugs inhibit p-glycoprotein (MDR1; ABCB1),[44] ABCG2[45] and ATP-binding cassette, subfamily C, member 1 (MRP1; ABCC1).[46] However, there are conflicting reports on whether either of these two NNRTIs are ABCB1 substrates.[47,48] The nuclear receptor, the constitutive androstane receptor (CAR), correlates with CYP2B6[49] and CYP2A6[50] expression in liver, even in the absence of enzyme inducers and xenobiotics. Activators of CAR have also been shown to induce *UGT2B* genes in vivo[51] and so CAR appears to play a role in basal and inducible regulation of all the enzymes involved in EFV metabolism. An SNP in *CAR* (CAR C/C genotype) has been shown to be associated with EFV plasma concentrations[52] and early treatment discontinuation of EFV-containing regimens.[53]

Downloaded From Vet-eBooks.com

CYP2B6 is primarily expressed in the liver and is one of the most polymorphic CYP genes in humans. Its variants have shown to affect transcriptional regulation, splicing, mRNA and protein expression, and catalytic activity.[54] Three *CYP2B6* polymorphisms in particular (516G>T, 983T>C, and 15,582C>T) predict increased plasma EFV exposure. The various combinations of these loss-of-function alleles, which are common in all race/ethnicity, define 10 plasma EFV concentration strata spanning an about 10-fold range.[55,56] The impact of the *CYP2B6* 516G>T SNP on the pharmacokinetics and pharmacodynamics of EFV is well described.[57] Heterozygous and homozygous carriers of the variant allele have been shown to express up to fourfold less CYP2B6 protein in comparison to the individuals homozygous for the common allele, resulting in increased EFV levels. A number of studies have shown that the 516G>T SNP is also associated with NVP plasma concentrations.[58–61]

The *CYP2B6* 983 T>C SNP, predominantly found in West African populations, reduces expression of CYP2B6[62,63] and is associated with up to threefold higher plasma concentrations of EFV, which rises to up to fivefold if the *CYP2B6* 516 G>T SNP is also present.[62] Heterozygosity for the 983T>C SNP has also been reported to be associated with significantly higher NVP plasma levels in Black patients.[64] In a 2015 study, a novel haplotype of *CYP2B6* consisting of three polymorphisms (rs10403955, rs2279345, and rs8192719) was shown to be more strongly associated with EFV concentrations above the reported minimum toxic concentration (4000 ng/mL) than the 516G>T polymorphism.[65] Similarly, *CYP2B6* polymorphisms have been found to be associated with extremely high concentrations of EFV in Japanese patients receiving a standard dose allowing a dose reduction to 200 mg without loss of antiviral efficacy and improvement in CNS symptoms.[66,67] Importantly, efficacy was not associated with CYP2B6 polymorphisms within the Encore I study, even in patients with the highest genetic susceptibility receiving 400 mg once daily.[68,69]

In the presence of *CYP2B6* slow metabolizer genotypes, EFV clearance depends on secondary pathways such that CYP2A6 (−48T>G) and UGT2B7 (735A G/G homozygote) associate with even greater EFV concentrations. In a 2015 study, this was seen in all subjects, and in Black subjects and White subjects analyzed separately. It was also independent of *CYP2B6* slow metabolizer genotype subgroup (i.e., 516 TT, 516 T/983C, and 983CC).[70] There is a reported weak association between *CYP3A4* (−392A>G) and *CYP3A5* (6986A>G) variants and EFV exposure, which is diminished when the study population is stratified by ethnicity.[57,71] The Swiss HIV cohort have also demonstrated that in the presence of defective CYP2B6 metabolism, there is a significant association between EFV exposure and *CYP3A4* and *CYP2A6* variants.[72,73]

Higher plasma concentrations of EFV are associated with central nervous system side effects.[74–77] *CYP2B6* 516G>T is associated with EFV-induced CNS toxicity during the first week of therapy.[57] A composite *CYP2B6* 516/983 "slow metabolizer" genotype has also been shown to correlate with a higher rate of CNS side effects in Caucasians,[78] and *CYP2B6* polymorphisms influence EFV (but not metabolite) concentration in cerebrospinal fluid.[79] A number of studies have attempted to individualize EFV dosing using prospective *CYP2B6* 516G>T genotyping,[66,80,81] some showing that *CYP2B6* genotyping can decrease CNS side effects and can also reduce

Downloaded From Vet-eBooks.com

treatment costs.[55,72] A genetic association study involving over 1000 AIDS Clinical trials (ACTG) participants suggested that most patients could receive less than 600 mg doses of EFV.[56]

A 2014 study showed an association with the *CYP2B6* 983 T>C SNP- and NVP-induced Steven Johnson syndrome but no other skin hypersensitivity in Malawian and Ugandan HIV-infected individuals.[82] As *CYP2B6* 983T>C has a frequency of 5−10% in a variety of African populations, but is not observed in Caucasians, this represents an ethnic-specific predisposing factor; it also highlights that genetic-susceptibility markers differ for NVP liver events and skin events without liver involvement.[83] A retrospective study on a pediatric cohort on NVP-based regimens demonstrated that the percentage increase in CD4 cell count was 3 times higher in patients with the *CYP2B6* 516 TT genotype compared to those with the 516 GG genotype.[61]

In one study, HLA-DRB1*0101 predicted the development of NVP hepatotoxicity in a cohort consisting of predominantly Caucasian patients with CD4 cell percentages of greater than 25%.[84] In this study, the occurrence of an isolated rash was not associated with CD4 cell percentage or HLA-DRB1*0101. In another study, the occurrence of an isolated rash in Caucasian patients on EFV and NVP-based regimens was associated with the presence of HLA-DRB1*0101 but was not associated with CD4 cell percentages.[85] In the latter, 83% of the participants presenting with isolated rashes were HLA-DRB1*0101 positive as compared to the 7% in the tolerant group. HLA-cw8 has been reported to be a significant predictor of NVP HSR in Sardinian and Japanese populations.[86,87] Additionally, despite the lack of a defined role for ABCB1 in the disposition of NNRTIs, *ABCB1* 3435C>T has been reported to be associated with a decreased risk of NNRTI-associated hepatotoxicity.[88,89]

Finally, etravirine and rilpivirine are second-generation NNRTIs. The influence of several SNPs at *CYP3A4*, *CYP2C9*, *CYP2C19*, and *UGT* genes on etravirine plasma levels was examined in 2013; only the *CYP2C19*2* haplotype caused a 25% reduction in the hepatic clearance of the drug, which is considered not clinically relevant.[90] No published studies have yet robustly assessed pharmacogenetic determinants of rilpivirine pharmacokinetics and response.

2.3 Protease Inhibitors

Before the development of newer drug classes, PIs represented the "last option" and were often used in various combinations to offer patients salvage therapy. Most of them caused considerable gastrointestinal symptoms and are associated with a metabolic syndrome including dyslipidemia, impaired insulin resistance, and lypodystrophy.[91] Patients receiving either ATV or indinavir have an increased risk of developing unconjugated bilirubinemia. Indinavir therapy is associated with the development of renal calculi while tipranavir and darunavir can be associated with significant hepatotoxicity.[92] However, with more favorable toxicity and tolerability profiles and once-daily dosing, darunavir and ATV are the main PIs prescribed today. LPV remains second-line therapy in developing countries.[34]

PIs are principally metabolized by CYP3A enzymes[93] and are normally administered with a potent CYP3A4 inhibitor (either a low dose of ritonavir or more recently

Downloaded From Vet-eBooks.com

cobicistat) to improve their bioavailability.[94] They are highly protein bound to both albumin and α1-acid glycoprotein (AAG; orosomucoid; ORM1)[95] and are ABCB1 substrates.[96] They have also been shown to be substrates for influx transporters OATP1A2, OATP1B1, and OATP1B3.[97]

CYP3A5 expressors (defined as individuals with the A allele for the *CYP3A5* 6986A>G polymorphism) have been reported to show faster oral clearance of indinavir[14] and saquinavir.[98–100] Nelfinavir is predominantly metabolized by CYP2C19 into its major metabolite M8 with a minor contribution from CYP3A4.[101] There is a confirmed association between *CYP2C19* 681G>A and higher nelfinavir exposure.[102,103] However, since these PIs are no longer used frequently, these data are of more mechanistic than clinical value. ABCB1 is thought to be an important efflux transporter in expelling PIs from the cell. The most studied variant, *ABCB1* 3435C>T is a synonymous SNP, which is believed to change substrate specificity.[104] It has been associated with increased ATV plasma levels and hyperbilirubinemia in Spanish patients.[105] In 2014, the intracellular/plasma concentration ratio of ATV was found to be higher in GG carriers compared with those with GT and TT genotypes of the *ABCB1* 2677 G>T SNP in an Italian study.[106] However, the impact of *ABCB1* variants on the expression and function of ABCB1, and the pharmacokinetics of its substrates remain controversial.

It has been speculated that the high protein binding associated with PIs may also contribute to the intraindividual variability of PI disposition. AAG variants have been reported to increase the apparent clearance of both LPV and indinavir to varying degrees, without impacting on the cellular exposure of either drug.[107] The significance of these findings remains unclear. Several groups have also associated *SLCO1B1* (coding for OATP1B1) polymorphisms with LPV plasma concentrations,[62,97,104,108,109] and the 521T>C polymorphism, while rare, appears to have a profound impact on the drug exposure.[110] A 2014 study showed the combined influence of *SLCO1B1* polymorphisms with a relatively recently characterized metabolic variant, CYP3A4*22.[111] In a 2014 study it was found that polymorphisms in *SLCO1B1* was associated with pharmacokinetic exposure to darunavir.[112] Finally, a pregnane X receptor (PXR) polymorphism was associated with plasma concentrations of unboosted ATV.[113,114] Since PXR influences the expression of ABCB1 and CYP3A4 in the liver (even in the absence of enzyme inducers), there is a biologically plausible mechanism for this association.[115,116]

ATV inhibits UDP-glucuronosyltransferase 1A1 (UGT1A1)–mediated bilirubin glucoronidation and 20–50% of patients develop unconjugated hyperbilirubinemia to a moderate degree while taking ATV, through a mechanism that mimics that of Gilbert's syndrome.[117] A few patients, however, develop overt, stigmatizing jaundice sufficient to consider discontinuation of therapy (6%).[117] A promoter polymorphism in UGT1A1 (UGT1A1*28) predicts reduced UGT1A1 expression and is associated with the occurrence of unconjugated hyperbilirubinemia in Swiss and Spanish patients on ATV and indinavir.[118–120] In contrast with Caucasians, the 211G>A polymorphism (*UGT1A1*6*) is a better predictor of jaundice in Asian patients treated with ATV.[120] Genotyping for *UGT1A1* and for *ABCB1* (notwithstanding the previously mentioned caveats for the latter) may help identify patients at high risk of hyperbilirubinemia.[121]

Downloaded From Vet-eBooks.com

Therapy containing boosted PIs has long been associated with dyslipidemia and metabolic disorders. Genetic predispositions include the ABCA1 296A>G SNP, which has been associated with increased HDL-cholesterol plasma levels after ritonavir boosted PI therapy in the Swiss HIV cohort.[122] The contribution of other SNPs associated with dyslipidemia in the general non-HIV-infected population has also been studied. SNPs in polymorphic lipid transport proteins Apolipoprotein A5 (APOA5 -1131T> and 64G>C), APOE E (three major isoforms: apo ε2, apo ε3, and apo ε4), and APOC3 (−482C>T, 455T>C, 3238C>G) have been shown to contribute to increased plasma triglycerides, HDL cholesterol, and/or LDL cholesterol during ART.[122−127] The Swiss HIV cohort used scoring algorithms to correlate the degree of hyperlipidemia with the number of unfavorable polymorphisms. In one study, individuals with unfavorable APOE isoforms (ε2 or ε4) as well as more than two of the APOC3 variants were observed to have significant hypertriglyceridemia (>6 mmol/L) if they received ritonavir-containing antiretroviral regimens.[123] In a subsequent study, the same group added APOA5 (non*1/*1 haplotypes), ABC transporter A1 (*ABCA1*; 2962A>G), and cholesteryl ester transfer protein (*CETP*; 279G>A) variants to create a scoring algorithm containing the *ABCA1/APOA5/APOC3/APOE/CETP* genotype composite score and the type of antiretroviral therapy patients were on.[122] Both longitudinal studies were performed in predominantly Caucasian cohorts and therefore validation studies should also explore the associations in other ethnicities.

2.4 Entry and Integrase Inhibitors

Maraviroc is a CCR5 antagonist, which is metabolized by CYP3A4 and CYP3A5.[128] Therefore, many of the associations described earlier for PIs may be relevant to this drug also. Indeed, hepatic uptake of Maraviroc is influenced by the action of the carrier protein OATP1B1, encoded by the *SLCO1B1* gene. Moreover, 521T>C polymorphisms in *SLCO1B1* is correlated with higher plasma MVC concentrations.[121] In addition, a 2014 work indicated a role of *CYP3A5* polymorphisms in influencing MVC exposure,[129] although the clinical relevance of this has been debated.[130,131]

Raltegravir (RAL), Elvitegravir (EVG), and Dolutegravir (DTG) are among the latest drugs introduced into the HIV armamentarium. RAL is predominantly metabolized by UGT1A1.[132] *UGT1A1*28* increases RAL concentrations but with little consequence, given its wide therapeutic range and intrinsic potency.[133−135] RAL diffusion into the cerebral spinal fluid was not affected by variability at genes *ABCB1*, *ABCC2*, *SLC22A2*, and/or *SLC22A6* that encode for drug transporters.[135,136] The rGC for RAL was shown to be low[2] but this may be heavily influenced for this drug by variability as a result of issues with intestinal pH, tablet dissolution, and interactions with antacids.[137,138]

DTG is also conjugated by UGT1A1 and homozygosity for the same UGT1A1 promoter variant correlates with about 50% greater plasma DTG concentration,[139] which has been judged not to be clinically significant. However, such information might be useful in patients receiving concomitant medications that increase (e.g., ATV) or decrease (e.g., DRV) dolutegravir exposure, or when underlying integrase inhibitor resistance suggests the need for higher daily dose.[83]

Downloaded From Vet-eBooks.com

3. Pharmacogenetics of Antimalarial Therapy

3.1 Artemisinin Compounds

Artemisinin-based combination therapies (ACTs) are recommended by WHO as the first-line treatment for uncomplicated *Plasmodium falciparum* malaria.

The metabolism of artemisinin derivatives is complex. Artesunate, artemether, and arteether are primarily metabolized by CYP3A4, CYP3A5, and CYP2A6 with a minor contribution from CYP2B6 to form dihydroartemisinin, the active compound.[140] Dihydroartemisinin is subsequently inactivated via UGT1A9 and UGT2B7.[141] Artemisinin on the other hand is primarily metabolized by CYP2B6 with a minor contribution from CYP3A4 and CYP2A6.[142,143] Artemether, artemisinin, arteether, and dihydroartemisinin have all been shown to induce CYP3A4, CYP2B6, and ABCB1 through activation of PXR and constitutive androstane receptor (CAR).[143–145] Despite the increasing call for pharmacogenetic-guided policies in the treatment of malaria in endemic areas, pharmacogenetic data with clinical endpoints on artemesinin compounds are still sparse.[146] However, associations described with antiretroviral compounds may also be relevant given the similarity in drug disposition pathways.

For instance, CYP2A6 is a major metabolizer of artesunate (AS) to its active metabolite, while CYP2B6 plays a more minor role.[147] The *CYP2A6* slow metabolizer status is associated with treatment failure and is thought to contribute to apparent "artemisinin resistance" in southeast Asia.[148] Conversely, in a 2012 study, Malaysian subjects carrying the *CYP2A6*1B* variants, responsible for ultrarapid metabolism of AS, suffered a significantly higher incidence of adverse events, secondary to accumulation of the active metabolites.[149] Finally, in a study of Burmese patients, the proportion of individuals with adequate clinical and parasitological response who had the *CYP2B6* 516G>T genotype (poor metabolizer genotype) was significantly lower compared with those with late parasitological failure (14.0% versus 19.0%).[150] These findings were, however, not reproduced in a study involving Cambodians and Tanzanians.[151] *CYP2B6* 516G>T has also been associated with raised plasma concentrations of artemesinin and artmether.[147,152] The *CYP3A4*1B* variant allele, which is associated with higher expression of CYP3A4, is associated with poor metabolism of artemether, lumefantrine.[145,151] Finally, CYP3A5*3 has also been suggested as a poor metabolizer allele for artemether, arteether, and artemisinin.[147,153] Overall, however, pharmacodynamic data with parasitocidal activity and toxicity correlations for these associations is still lacking.

Lumefantrin is an importing partner drug in ACT and is administered exclusively with artemether for the treatment of uncomplicated malaria by *P. falciparum*. It is mainly metabolized by CYP3A4 to the pharmacologically active desbutyllumefantrine and has very variable bioavailability. A 2016 study reported that due to long-term CYP3A induction, EFV-based ART cotreatment significantly reduces lumefantrine plasma exposure leading to poor malaria treatment response in HIV coinfected individuals, this was more pronounced in *CYP2B6* slow metabolizers (516 G>T). In this study, NVP coadministration did not show decreased lumefantrine levels, which may reflect its milder effect on CYP3A4 induction.[154] The authors highlighted the

Downloaded From Vet-eBooks.com

importance of pharmacogenomics in the management of malaria and HIV cotreatment, particularly in resource-poor settings where the main burden for both diseases stands and where EFV is first-line ARV treatment.

3.2 Primaquine

Primaquine is used in patients with *Plasmodium ovale* and *Plasmodium vivax* infections to clear the latent hepatic hypnozoite stage of the parasite and prevent relapse and transmissibility. It is primarily metabolized by CYP1A2 and CYP3A4.[155] Although ethnicity has been found to be significantly associated with plasma levels of primaquine, the genetic polymorphisms underpinning this variability, or their impact on efficacy or toxicity are yet to be characterized.[39,156,157]

The first records of variability in response to antimalarials dates back to World War II when African−American soldiers were found to experience higher rates of acute hemolysis when they received primaquine compared to their Caucasian counterparts.[158] The basis of these observed differences was later attributed to glucose-6-phosphate dehydrogenase (G6PD) deficiency, an X-linked recessive disorder,[159] common in sub-Saharan Africans (∼10−25%).[160] The *G6PD* locus is highly polymorphic, so in clinical practice, prospective qualitative and quantitative tests are performed in patients requiring primaquine.[161]

3.3 Amodiaquine

Amodiaquine, combined with artesunate, is commonly used in Africa. It is primarily metabolized by CYP2C8 into its active form[162] and exhibits huge interindividual variability in plasma drug concentrations.[163,164] More specifically, the *CYP2C8*2* and *CYP2C8*3* alleles have been associated with a reduction in amodiaquine metabolism (*CYP2C8*3* most significantly). Toxic metabolites (quinoneimines; QNMs) are more likely to be formed in this setting of *CYP2C8* slow metabolizer genotypes,[147] potentially exacerbating agranulocytosis and severe liver damage. In late 2000s, in vivo studies have shown that extrahepatic metabolism by CYP1A1 and CYP1B1 may also generate QNMs[165,166] but no genetic data on their variability are currently available. Of note, *CYP2C8*2* is the variant most common in African populations while *CYP2C8*3* is the variant most common in Caucasian populations.[167,168] In early 2010s an association between host *CYP2C8*3* carriage and parasitological relapse with *P. falciparum* carrying the amodiaquine resistance (via *pfmdr1*) was reported,[169,170] providing an example of how host genetic variation may influence the selection dynamics of a pathogen's drug resistance.

3.4 Mefloquine

Mefloquine is primarily metabolized by CYP3A4[171] to pharmacologically inactive metabolites and is suspected to be a ABCB1 substrate.[172] It is used in chemoprophylaxis and as accompanying agent to artenusate in ACT; the most common side effects are dose-dependent neuropsychiatric effects. The *ABCB1* 1236 TT/2677 TT/3455 TT

Downloaded From Vet-eBooks.com

haplotype has been reported to be associated with increased neuropsychiatric events in a homogenous Caucasian cohort, which were not related to mefloquine serum concentrations.[173] It is thought that a lower expression of ABCB1 in individuals carrying variants[174] results in lower mefloquine efflux from the brain, exposing individuals to high tissue concentrations related to neuropsychiatric symptoms. This finding might suggest the important role of local ABCB1 expression at the blood−brain barrier that leads to the CNS accumulation of mefloquine without affecting systemic exposure.[147] However, as discussed earlier, ABCB1 associations are controversial and it is difficult to understand why a systemic effect would not be observed. A 2015 study identified carboxymefloquine, the major and pharmacologically inactive metabolite of mefloquine, as a PXR activator in vitro leading, in turn, to the induction and expression of drug-metabolizing enzymes and transporters at the mRNA and protein levels, with potential impact drug−drug interactions.[175]

3.5 Proguanil

Proguanil is a component of the widely prescribed malarone (atovaquone-proguanil), which is used in both chemoprophylaxis and in treatment of malaria. The bioactivation of proguanil is primarily through CYP2C19 with a minor role from CYP3A4. *CYP2C19* variants have been associated with proguanil plasma concentrations[176,177]; for instance in a study in Gambian adults, ultrarapid metabolizers (*CYP2C19*17* homozygotes) had higher AUC and Cmax values for the active metabolites.[178] However, other studies have shown no correlation with clinical outcomes or adverse events.[147,179,180]

3.6 Quinine

Despite good efficacy,[181] quinine is no longer a WHO-favored agent for the treatment of *P. falciparum* infection, secondary to its poor toxicity and tolerability profiles.[182] Adverse events, if present, include prolonged QT interval, hypoglycemia, cinchonism, tinnitus, and vomiting.[183] Quinine is metabolized by CYP3A isoforms to its primary metabolite, 3-hydroxyquinine.[184,185] A study comparing the impact of *CYP3A5* genotypes on the hydroxylation of quinine between Tanzanians and Swedes found lower hydroxylation in Tanzanians that were homozygous for *CYP3A5*3*.[185] This finding is yet to be confirmed in other populations. Other potential associations include *ABCB1* polymorphisms with quinine neurotoxicity and OCT-2-related pancreatic insulin secretion in quinine-induced hypoglycemia.[145,147]

4. Pharmacogenetics of Antituberculous Therapy

Resistance is a big issue in tuberculosis (TB), especially with the emergence of multidrug-resistant TB (MDR TB) and more recently, in 2016, extensively drug-resistant TB (XDR TB). The unprecedented reemergence of TB with the HIV pandemic complicates matters further. The drugs used in first-line therapy have

Downloaded From Vet-eBooks.com

remained unchanged for over half a century. A number of newer agents are now in development and the challenges for clinical assessment were reviewed.[186]

During the early 1950s, individuals receiving isoniazid (INH) for the treatment of TB were noted to have marked differences in the amount of isoniazid excreted in their urine.[187] The basis of these differences was later attributed to differences in an individual's ability to acetylate isoniazid[188] via arylamine N-acetyltransferase 2 (NAT2).[189,190] The status of NAT2 activity is genetically controlled and depends on the number of alleles. Variant alleles (*NAT2*5*, **6*, **7*, **14*, and **19*) produce impaired NAT2 enzyme with lower activity.[190] Isoniazid toxicity mainly manifests itself as peripheral neuropathy and hepatotoxicity.[191,192] Slow acetylators are prone to isoniazid-related adverse events and, conversely, treatment failure is more likely in rapid acetylators (i.e., two active alleles), which are more frequent in Asian than in Caucasian populations.[193] While pharmacogenetics-based therapy models have been successfully piloted in some populations,[190,194] *NAT2* testing is not currently done in routine clinical practice. To prevent the development of peripheral neuropathy, however, pyridoxine is prescribed, alongside isoniazid, in all patients.

Hepatotoxicity is the most frequent side effect of first-line antituberculous compounds (1−33%).[195] A metaanalysis looking at reported associations between antituberculous drug-induced liver injury (ADLI) and drug-metabolizing variants identified homozygotes for variants of *NAT2, CYP2E1*, and *GSTM1* as significant predictors of hepatotoxicty.[196] However, it is worth noting that most of these studies were performed in Asian populations, on varying anti-TB medications with unstandardized definitions of hepatotoxicity and uncharacterized environmental factors. Indeed, studies in early 2010s in diverse populations have not always replicated these results; for instance, there was no association between increased risk of ADLI and the presence of slow *NAT2* polymorphisms in Caucasian patients in a case control study[197] but there was with a *GSTT1* homozygous null genotype in an earlier study.[198] In other studies in Indian[199,200] and Korean[200] populations, neither *GSTM1* nor *GSTT1* null genotypes were associated with anti-TB drug-induced hepatotoxicity. Further clarification is therefore needed.

Finally, rifampicin is a substrate of drug transporters, such as ABCB1 and OATP1B1, which are transcriptionally regulated by nuclear receptors, such as PXR and CAR.[189] In one study, the pharmacokinetics of rifampicin were shown to be associated with a polymorphism within the *SLCO1B1* gene.[201] In this report, the rifampicin AUC was about 36% lower in patients with the *SLCO1B1 463CA* genotype compared to patients homozygous for the C allele. Importantly, *SLCO1B1* polymorphisms associated with lower rifampicin exposure were more frequent in Black subjects. Additionally, in a study in South Africa, patients heterozygous and homozygous for the variant allele of *SLCO1B1* (rs4149032) polymorphism had lower rifampicin bioavailability of 20% and 28%, respectively. Simulations revealed that an increase in rifampicin dose of about 30% in patients harboring the polymorphism would result in plasma rifampicin levels similar to those in noncarriers. Other polymorphisms in *ABCB1*, *PXR*, and *CAR* did not exhibit any significant impact on the pharmacokinetics of rifampicin.[202] Limited studies are available for newer rifamycins but a 2015 study indicated that *SLCO1B1* polymorphisms may also influence rifabutin pharmacokinetics.[203] In 2015, the pharmacogenetics of tuberculosis therapy and special populations was reviewed.[204]

Downloaded From Vet-eBooks.com

5. Summary and Perspective

In infectious diseases, pharmacogenetic testing should ideally be implemented in conjunction with pathogen resistance testing and the characterization of a compound's pharmacokinetic and pharmacodynamic attributes. Apart from the idiosyncratic HSRs seen with ABC and NVP, most studied clinical phenotypes are much more subtle, develop over time, and are multifactorial in etiology. Current general challenges within infectious diseases include antimicrobial resistance and limited effective therapies for emerging, reemerging, and neglected infections. Pharmacogenetic studies of both human host and disease pathogens may help tackle the resistance issue, and genome-wide studies to identify new drug targets for emerging, reemerging, and neglected infections are warranted.

References

1. Kalow W, Tang BK, Endrenyi L. Hypothesis: comparisons of inter- and intra-individual variations can substitute for twin studies in drug research. *Pharmacogenetics* 1998;**8**: 283–9.
2. Siccardi M, Olagunju A, Simiele M, et al. Class-specific relative genetic contribution for key antiretroviral drugs. *J Antimicrob Chemother* 2015;**70**:3074–9.
3. Setzer B, Schlesier M, Thomas AK, Walker UA. Mitochondrial toxicity of nucleoside analogues in primary human lymphocytes. *Antivir Ther* 2005;**10**:327–34.
4. Brinkman K, Smeitink JA, Romijn JA, Reiss P. Mitochondrial toxicity induced by nucleoside-analogue reverse-transcriptase inhibitors is a key factor in the pathogenesis of antiretroviral-therapy-related lipodystrophy. *Lancet* 1999;**354**:1112–5.
5. Group DADS, Sabin CA, Worm SW, et al. Use of nucleoside reverse transcriptase inhibitors and risk of myocardial infarction in HIV-infected patients enrolled in the D:A:D study: a multi-cohort collaboration. *Lancet* 2008;**371**:1417–26.
6. Marcus JL, Neugebauer RS, Leyden WA, et al. Use of abacavir and risk of cardiovascular disease among HIV-infected individuals. *J Acquir Immune Defic Syndr* 2015;**69**: 413.
7. Calmy A, Hirschel B, Cooper DA, Carr A. Clinical update: adverse effects of antiretroviral therapy. *Lancet* 2007;**370**:12–4.
8. Stellbrink HJ, Orkin C, Arribas JR, et al. Comparison of changes in bone density and turnover with abacavir-lamivudine versus tenofovir-emtricitabine in HIV-infected adults: 48-week results from the ASSERT study. *Clin Infect Dis* 2010;**51**:963–72.
9. Wohl D, Oka S, Clumeck N, et al. A randomized, double-blind comparison of tenofovir alafenamide (TAF) vs. Tenofovir disoproxil fumarate (TDF), each coformulated with Elvitegravir, cobicistat, and emtricitabine (E/C/F) for initial HIV-1 treatment: week 96 results. *J Acquir Immune Defic Syndr* 2016;**72**(1):58–64.
10. Veal GJ, Back DJ. Metabolism of zidovudine. *Gen Pharmacol* 1995;**26**:1469–75.
11. Schuetz JD, Connelly MC, Sun D, et al. MRP4: a previously unidentified factor in resistance to nucleoside-based antiviral drugs. *Nat Med* 1999;**5**:1048–51.
12. Takenaka K, Morgan JA, Scheffer GL, et al. Substrate overlap between Mrp4 and Abcg2/Bcrp affects purine analogue drug cytotoxicity and tissue distribution. *Cancer Res* 2007; **67**:6965–72.

Downloaded From Vet-eBooks.com

13. Wijnholds J, Mol CA, van Deemter L, et al. Multidrug-resistance protein 5 is a multispecific organic anion transporter able to transport nucleotide analogs. *Proc Natl Acad Sci USA* 2000;**97**:7476−81.
14. Anderson PL, Lamba J, Aquilante CL, Schuetz E, Fletcher CV. Pharmacogenetic characteristics of indinavir, zidovudine, and lamivudine therapy in HIV-infected adults: a pilot study. *J Acquir Immune Defic Syndr* 2006;**42**:441−9.
15. Mallal S, Nolan D, Witt C, et al. Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. *Lancet* 2002;**359**:727−32.
16. Mallal S, Phillips E, Carosi G, et al. HLA-B*5701 screening for hypersensitivity to abacavir. *N Engl J Med* 2008;**358**:568−79.
17. Hughes AR, Mosteller M, Bansal AT, et al. Association of genetic variations in HLA-B region with hypersensitivity to abacavir in some, but not all, populations. *Pharmacogenomics* 2004;**5**:203−11.
18. Cutrell AG, Hernandez JE, Fleming JW, et al. Updated clinical risk factor analysis of suspected hypersensitivity reactions to abacavir. *Ann Pharmacother* 2004;**38**:2171−2.
19. Saag M, Balu R, Phillips E, et al. High sensitivity of human leukocyte antigen-b*5701 as a marker for immunologically confirmed abacavir hypersensitivity in white and black patients. *Clin Infect Dis* 2008;**46**:1111−8.
20. *Allele frequencies in worldwide populations*; 2014. Available from http://www.allelefrequencies.net/default.asp
21. Uwai Y, Ida H, Tsuji Y, Katsura T, Inui K. Renal transport of adefovir, cidofovir, and tenofovir by SLC22A family members (hOAT1, hOAT3, and hOCT2). *Pharm Res* 2007;**24**:811−5.
22. Ray AS, Cihlar T, Robinson KL, et al. Mechanism of active renal tubular efflux of tenofovir. *Antimicrob Agents Chemother* 2006;**50**:3297−304.
23. Moss DM, Neary M, Owen A. The role of drug transporters in the kidney: lessons from tenofovir. *Front Pharmacol* 2014;**5**:248.
24. Rodriguez-Novoa S, Labarga P, Soriano V, et al. Predictors of kidney tubular dysfunction in HIV-infected patients treated with tenofovir: a pharmacogenetic study. *Clin Infect Dis* 2009;**48**:e108−16.
25. Nishijima T, Komatsu H, Higasa K, et al. Single nucleotide polymorphisms in ABCC2 associate with tenofovir-induced kidney tubular dysfunction in Japanese patients with HIV-1 infection: a pharmacogenetic study. *Clin Infect Dis* 2012;**55**:1558−67.
26. Sirirungsi W, Urien S, Harrison L, et al. No relationship between drug transporter genetic variants and tenofovir plasma concentrations or changes in glomerular filtration rate in HIV-infected adults. *J Acquir Immune Defic Syndr* 2015;**68**:e56−9.
27. Pushpakom SP, Liptrott NJ, Rodriguez-Novoa S, et al. Genetic variants of ABCC10, a novel tenofovir transporter, are associated with kidney tubular dysfunction. *J Infect Dis* 2011;**204**:145−53.
28. Calcagno A, Cusato J, Marinaro L, et al. Clinical pharmacology of tenofovir clearance: a pharmacokinetic/pharmacogenetic study on plasma and urines. *Pharmacogenomics J* 2015.
29. Giacomet V, Cattaneo D, Vigano A, et al. Tenofovir-induced renal tubular dysfunction in vertically HIV-infected patients associated with polymorphisms in ABCC2, ABCC4 and ABCC10 genes. *Pediatr Infect Dis J* 2013;**32**:e403−5.
30. Gazzard BG, Anderson J, Babiker A, et al. British HIV Association Guidelines for the treatment of HIV-1-infected adults with antiretroviral therapy 2008. *HIV Med* 2008;**9**:563−608.

Downloaded From Vet-eBooks.com

31. Clumeck N, Pozniak A, Raffi F, Committee EE. European AIDS Clinical Society (EACS) guidelines for the clinical management and treatment of HIV-infected adults. *HIV Med* 2008;**9**:65−71.
32. EACS. *Guidelines V8.0 October 2015*. 2015.
33. DHHS. *Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents*. 2016.
34. *Antiretroviral therapy for HIV infection in adults and adolescents: recommendations for a public health approach - 2010 revision*. 2010.
35. Wainberg MA. HIV resistance to nevirapine and other non-nucleoside reverse transcriptase inhibitors. *J Acquir Immune Defic Syndr* 2003;**34**(Suppl. 1):S2−7.
36. Stern JO, Robinson PA, Love J, Lanes S, Imperiale MS, Mayers DL. A comprehensive hepatic safety analysis of nevirapine in different populations of HIV infected patients. *J Acquir Immune Defic Syndr* 2003;**34**(Suppl. 1):S21−33.
37. Dieterich DT, Robinson PA, Love J, Stern JO. Drug-induced liver injury associated with the use of nonnucleoside reverse-transcriptase inhibitors. *Clin Infect Dis* 2004;**38**(Suppl. 2):S80−9.
38. Gazzard BG. Efavirenz in the management of HIV infection. *Int J Clin Pract* 1999;**53**: 60−4.
39. Lochet P, Peyriere H, Lotthe A, Mauboussin JM, Delmas B, Reynes J. Long-term assessment of neuropsychiatric adverse reactions associated with efavirenz. *HIV Med* 2003;**4**:62−6.
40. Group ES, Carey D, Puls R, et al. Efficacy and safety of efavirenz 400 mg daily versus 600 mg daily: 96-week data from the randomised, double-blind, placebo-controlled, non-inferiority ENCORE1 study. *Lancet Infect Dis* 2015;**15**:793−802.
41. Erickson DA, Mather G, Trager WF, Levy RH, Keirns JJ. Characterization of the in vitro biotransformation of the HIV-1 reverse transcriptase inhibitor nevirapine by human hepatic cytochromes P-450. *Drug Metab Dispos* 1999;**27**:1488−95.
42. Ward BA, Gorski JC, Jones DR, Hall SD, Flockhart DA, Desta Z. The cytochrome P450 2B6 (CYP2B6) is the main catalyst of efavirenz primary and secondary metabolism: implication for HIV/AIDS therapy and utility of efavirenz as a substrate marker of CYP2B6 catalytic activity. *J Pharmacol Exp Ther* 2003;**306**:287−300.
43. Kwara A, Lartey M, Sagoe KW, Kenu E, Court MH. CYP2B6, CYP2A6 and UGT2B7 genetic polymorphisms are predictors of efavirenz mid-dose concentration in HIV-infected patients. *AIDS* 2009;**23**:2101−6.
44. Storch CH, Theile D, Lindenmaier H, Haefeli WE, Weiss J. Comparison of the inhibitory activity of anti-HIV drugs on P-glycoprotein. *Biochem Pharmacol* 2007;**73**: 1573−81.
45. Weiss J, Rose J, Storch CH, et al. Modulation of human BCRP (ABCG2) activity by anti-HIV drugs. *J Antimicrob Chemother* 2007;**59**:238−45.
46. Weiss J, Theile D, Ketabi-Kiyanvash N, Lindenmaier H, Haefeli WE. Inhibition of MRP1/ABCC1, MRP2/ABCC2, and MRP3/ABCC3 by nucleoside, nucleotide, and non-nucleoside reverse transcriptase inhibitors. *Drug Metab Dispos* 2007;**35**: 340−4.
47. Stormer E, von Moltke LL, Perloff MD, Greenblatt DJ. Differential modulation of P-glycoprotein expression and activity by non-nucleoside HIV-1 reverse transcriptase inhibitors in cell culture. *Pharm Res* 2002;**19**:1038−45.
48. Almond LM, Hoggard PG, Edirisinghe D, Khoo SH, Back DJ. Intracellular and plasma pharmacokinetics of efavirenz in HIV-infected individuals. *J Antimicrob Chemother* 2005; **56**:738−44.

Downloaded From Vet-eBooks.com

49. Chang TK, Bandiera SM, Chen J. Constitutive androstane receptor and pregnane X receptor gene expression in human liver: interindividual variability and correlation with CYP2B6 mRNA levels. *Drug Metab Dispos* 2003;**31**:7−10.
50. Wortham M, Czerwinski M, He L, Parkinson A, Wan YJ. Expression of constitutive androstane receptor, hepatic nuclear factor 4 alpha, and P450 oxidoreductase genes determines interindividual variability in basal expression and activity of a broad scope of xenobiotic metabolism genes in the human liver. *Drug Metab Dispos* 2007;**35**:1700−10.
51. Shelby MK, Klaassen CD. Induction of rat UDP-glucuronosyltransferases in liver and duodenum by microsomal enzyme inducers that activate various transcriptional pathways. *Drug Metab Dispos* 2006;**34**:1772−8.
52. Cortes CP, Siccardi M, Chaikan A, Owen A, Zhang G, la Porte CJ. Correlates of efavirenz exposure in Chilean patients affected with human immunodeficiency virus reveals a novel association with a polymorphism in the constitutive androstane receptor. *Ther Drug Monit* 2013;**35**:78−83.
53. Wyen C, Hendra H, Siccardi M, et al. Cytochrome P450 2B6 (CYP2B6) and constitutive androstane receptor (CAR) polymorphisms are associated with early discontinuation of efavirenz-containing regimens. *J Antimicrob Chemother* 2011;**66**:2092−8.
54. Zanger UM, Klein K. Pharmacogenetics of cytochrome P450 2B6 (CYP2B6): advances on polymorphisms, mechanisms, and clinical relevance. *Front Genet* 2013;**4**:24.
55. Schackman BR, Haas DW, Park SS, Li XC, Freedberg KA. Cost-effectiveness of CYP2B6 genotyping to optimize efavirenz dosing in HIV clinical practice. *Pharmacogenomics* 2015;**16**:2007−18.
56. Holzinger ER, Grady B, Ritchie MD, et al. Genome-wide association study of plasma efavirenz pharmacokinetics in AIDS Clinical Trials Group protocols implicates several CYP2B6 variants. *Pharmacogenet Genomics* 2012;**22**:858−67.
57. Haas DW, Ribaudo HJ, Kim RB, et al. Pharmacogenetics of efavirenz and central nervous system side effects: an Adult AIDS Clinical Trials Group study. *AIDS* 2004;**18**:2391−400.
58. Mahungu T, Smith C, Turner F, et al. Cytochrome P450 2B6 516G→T is associated with plasma concentrations of nevirapine at both 200 mg twice daily and 400 mg once daily in an ethnically diverse population. *HIV Med* 2009;**10**:310−7.
59. Penzak SR, Kabuye G, Mugyenyi P, et al. Cytochrome P450 2B6 (CYP2B6) G516T influences nevirapine plasma concentrations in HIV-infected patients in Uganda. *HIV Med* 2007;**8**:86−91.
60. Rotger M, Colombo S, Furrer H, et al. Influence of CYP2B6 polymorphism on plasma and intracellular concentrations and toxicity of efavirenz and nevirapine in HIV-infected patients. *Pharmacogenet Genomics* 2005;**15**:1−5.
61. Saitoh A, Sarles E, Capparelli E, et al. CYP2B6 genetic variants are associated with nevirapine pharmacokinetics and clinical response in HIV-1-infected children. *AIDS* 2007; **21**:2191−9.
62. Wang J, Sonnerborg A, Rane A, et al. Identification of a novel specific CYP2B6 allele in Africans causing impaired metabolism of the HIV drug efavirenz. *Pharmacogenet Genomics* 2006;**16**:191−8.
63. Klein K, Lang T, Saussele T, et al. Genetic variability of CYP2B6 in populations of African and Asian origin: allele frequencies, novel functional variants, and possible implications for anti-HIV therapy with efavirenz. *Pharmacogenet Genomics* 2005;**15**: 861−73.
64. Wyen C, Hendra H, Vogel M, et al. Impact of CYP2B6 983T>C polymorphism on non-nucleoside reverse transcriptase inhibitor plasma concentrations in HIV-infected patients. *J Antimicrob Chemother* 2008;**61**:914−8.

Downloaded From Vet-eBooks.com

65. Carr DF, la Porte CJ, Pirmohamed M, Owen A, Cortes CP. Haplotype structure of CYP2B6 and association with plasma efavirenz concentrations in a Chilean HIV cohort. *J Antimicrob Chemother* 2010;**65**:1889−93.
66. Gatanaga H, Hayashida T, Tsuchiya K, et al. Successful efavirenz dose reduction in HIV type 1-infected individuals with cytochrome P450 2B6 *6 and *26. *Clin Infect Dis* 2007; **45**:1230−7.
67. Damronglerd P, Sukasem C, Thipmontree W, Puangpetch A, Kiertiburanakul S. A pharmacogenomic prospective randomized controlled trial of CYP2B6 polymorphisms and efavirenz dose adjustment among HIV-infected Thai patients: a pilot study. *Pharmgenomics Pers Med* 2015;**8**:155−62.
68. Dickinson L, Amin J, Else L, et al. Comprehensive pharmacokinetic, pharmacodynamic and pharmacogenetic evaluation of once-daily efavirenz 400 and 600 mg in treatment-naive HIV-infected patients at 96 weeks: results of the ENCORE1 study. *Clin Pharmacokinet* 2015;**15**(7):793−802.
69. Dickinson L, Amin J, Else L, et al. Pharmacokinetic and pharmacodynamic comparison of once-daily efavirenz (400 mg vs. 600 mg) in treatment-Naive HIV-infected patients: results of the ENCORE1 study. *Clin Pharmacol Ther* 2015;**98**:406−16.
70. David W, Haas AK, Richardson DM, et al. *Contribution of CYP2A6, UGT2B7, and other non-CYP2B6 polymorphisms to plasma efavirenz exposure*. Seatle: CROI; 2015.
71. Rodriguez-Novoa S, Barreiro P, Rendon A, Jimenez-Nacher I, Gonzalez-Lahoz J, Soriano V. Influence of 516G>T polymorphisms at the gene encoding the CYP450-2B6 isoenzyme on efavirenz plasma concentrations in HIV-infected subjects. *Clin Infect Dis* 2005;**40**:1358−61.
72. Martin AS, Gomez AI, Garcia-Berrocal B, et al. Dose reduction of efavirenz: an observational study describing cost-effectiveness, pharmacokinetics and pharmacogenetics. *Pharmacogenomics* 2014;**15**:997−1006.
73. di Iulio J, Fayet A, Arab-Alameddine M, et al. In vivo analysis of efavirenz metabolism in individuals with impaired CYP2A6 function. *Pharmacogenet Genomics* 2009;**19**:300−9.
74. Marzolini C, Telenti A, Decosterd LA, Greub G, Biollaz J, Buclin T. Efavirenz plasma levels can predict treatment failure and central nervous system side effects in HIV-1-infected patients. *AIDS* 2001;**15**:71−5.
75. Nunez M, Gonzalez de Requena D, Gallego L, Jimenez-Nacher I, Gonzalez-Lahoz J, Soriano V. Higher efavirenz plasma levels correlate with development of insomnia. *J Acquir Immune Defic Syndr* 2001;**28**:399−400.
76. Kappelhoff BS, van Leth F, Robinson PA, et al. Are adverse events of nevirapine and efavirenz related to plasma concentrations? *Antivir Ther* 2005;**10**:489−98.
77. Gallego L, Barreiro P, del Rio R, et al. Analyzing sleep abnormalities in HIV-infected patients treated with Efavirenz. *Clin Infect Dis* 2004;**38**:430−2.
78. Ribaudo HJ, Liu H, Schwab M, et al. Effect of CYP2B6, ABCB1, and CYP3A5 polymorphisms on efavirenz pharmacokinetics and treatment response: an AIDS Clinical Trials Group study. *J Infect Dis* 2010;**202**:717−22.
79. Winston A, Puls R. Cerebrospinal-fluid exposure of efavirenz and its major metabolites when dosed at 400 and 600 mg once daily; a randomized controlled trial. *J Int AIDS Soc* 2014;**17**:19541.
80. Rotger M, Telenti A. Optimizing efavirenz treatment: CYP2B6 genotyping or therapeutic drug monitoring? *Eur J Clin Pharmacol* 2008;**64**:335−6.
81. Torno MS, Witt MD, Saitoh A, Fletcher CV. Successful use of reduced-dose efavirenz in a patient with human immunodeficiency virus infection: case report and review of the literature. *Pharmacotherapy* 2008;**28**:782−7.

Downloaded From Vet-eBooks.com

82. Carr DF, Chaponda M, Cornejo Castro EM, et al. CYP2B6 c.983T>C polymorphism is associated with nevirapine hypersensitivity in Malawian and Ugandan HIV populations. *J Antimicrob Chemother* 2014;**69**:3329−34.
83. Haas DW, Tarr PE. Perspectives on pharmacogenomics of antiretroviral medications and HIV-associated comorbidities. *Curr Opin HIV AIDS* 2015;**10**:116−22.
84. Martin AM, Nolan D, James I, et al. Predisposition to nevirapine hypersensitivity associated with HLA-DRB1*0101 and abrogated by low CD4 T-cell counts. *AIDS* 2005;**19**: 97−9.
85. Vitezica ZG, Milpied B, Lonjou C, et al. HLA-DRB1*01 associated with cutaneous hypersensitivity induced by nevirapine and efavirenz. *AIDS* 2008;**22**:540−1.
86. Gatanaga H, Yazaki H, Tanuma J, et al. HLA-Cw8 primarily associated with hypersensitivity to nevirapine. *AIDS* 2007;**21**:264−5.
87. Littera R, Carcassi C, Masala A, et al. HLA-dependent hypersensitivity to nevirapine in Sardinian HIV patients. *AIDS* 2006;**20**:1621−6.
88. Haas DW, Bartlett JA, Andersen JW, et al. Pharmacogenetics of nevirapine-associated hepatotoxicity: an adult AIDS clinical trials group collaboration. *Clin Infect Dis* 2006; **43**:783−6.
89. Ritchie MD, Haas DW, Motsinger AA, et al. Drug transporter and metabolizing enzyme gene variants and nonnucleoside reverse-transcriptase inhibitor hepatotoxicity. *Clin Infect Dis* 2006;**43**:779−82.
90. Lubomirov R, Arab-Alameddine M, Rotger M, et al. Pharmacogenetics-based population pharmacokinetic analysis of etravirine in HIV-1 infected individuals. *Pharmacogenet Genomics* 2013;**23**:9−18.
91. Carr A, Samaras K, Burton S, et al. A syndrome of peripheral lipodystrophy, hyperlipidaemia and insulin resistance in patients receiving HIV protease inhibitors. *AIDS* 1998; **12**:F51−8.
92. Hughes CA, Robinson L, Tseng A, MacArthur RD. New antiretroviral drugs: a review of the efficacy, safety, pharmacokinetics, and resistance profile of tipranavir, darunavir, etravirine, rilpivirine, maraviroc, and raltegravir. *Expert Opin Pharmacother* 2009;**10**:2445−66.
93. Ernest 2nd CS, Hall SD, Jones DR. Mechanism-based inactivation of CYP3A by HIV protease inhibitors. *J Pharmacol Exp Ther* 2005;**312**:583−91.
94. Cooper CL, van Heeswijk RP, Gallicano K, Cameron DW. A review of low-dose ritonavir in protease inhibitor combination therapy. *Clin Infect Dis* 2003;**36**:1585−92.
95. Boffito M, Back DJ, Hoggard PG, et al. Intra-individual variability in lopinavir plasma trough concentrations supports therapeutic drug monitoring. *AIDS* 2003;**17**:1107−8.
96. Marzolini C, Paus E, Buclin T, Kim RB. Polymorphisms in human MDR1 (P-glycoprotein): recent advances and clinical relevance. *Clin Pharmacol Ther* 2004;**75**:13−33.
97. Hartkoorn RC, Kwan WS, Shallcross V, et al. HIV protease inhibitors are substrates for OATP1A2, OATP1B1 and OATP1B3 and lopinavir plasma concentrations are influenced by SLCO1B1 polymorphisms. *Pharmacogenet Genomics* 2010;**20**:112−20.
98. Mouly SJ, Matheny C, Paine MF, et al. Variation in oral clearance of saquinavir is predicted by CYP3A5*1 genotype but not by enterocyte content of cytochrome P450 3A5. *Clin Pharmacol Ther* 2005;**78**:605−18.
99. Frohlich M, Hoffmann MM, Burhenne J, Mikus G, Weiss J, Haefeli WE. Association of the CYP3A5 A6986G (CYP3A5*3) polymorphism with saquinavir pharmacokinetics. *Br J Clin Pharmacol* 2004;**58**:443−4.
100. Josephson F, Allqvist A, Janabi M, et al. CYP3A5 genotype has an impact on the metabolism of the HIV protease inhibitor saquinavir. *Clin Pharmacol Ther* 2007;**81**: 708−12.

Downloaded From Vet-eBooks.com

101. Zhang KE, Wu E, Patick AK, et al. Circulating metabolites of the human immunodeficiency virus protease inhibitor nelfinavir in humans: structural identification, levels in plasma, and antiviral activities. *Antimicrob Agents Chemother* 2001;**45**:1086–93.
102. Haas DW, Smeaton LM, Shafer RW, et al. Pharmacogenetics of long-term responses to antiretroviral regimens containing efavirenz and/or nelfinavir: an adult aids clinical trials group study. *J Infect Dis* 2005;**192**:1931–42.
103. Burger DM, Schwietert HR, Colbers EP, Becker M. The effect of the CYP2C19*2 heterozygote genotype on the pharmacokinetics of nelfinavir. *Br J Clin Pharmacol* 2006;**62**: 250–2.
104. Kimchi-Sarfaty C, Oh JM, Kim IW, et al. A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science* 2007;**315**:525–8.
105. Rodriguez Novoa S, Barreiro P, Rendon A, et al. Plasma levels of atazanavir and the risk of hyperbilirubinemia are predicted by the 3435C→T polymorphism at the multidrug resistance gene 1. *Clin Infect Dis* 2006;**42**:291–5.
106. D'Avolio A, Carcieri C, Cusato J, et al. Intracellular accumulation of atazanavir/ritonavir according to plasma concentrations and OATP1B1, ABCB1 and PXR genetic polymorphisms. *J Antimicrob Chemother* 2014;**69**:3061–6.
107. Colombo S, Buclin T, Decosterd LA, et al. Orosomucoid (alpha1-acid glycoprotein) plasma concentration and genetic variants: effects on human immunodeficiency virus protease inhibitor clearance and cellular accumulation. *Clin Pharmacol Ther* 2006;**80**: 307–18.
108. Lubomirov R, di Iulio J, Fayet A, et al. ADME pharmacogenetics: investigation of the pharmacokinetics of the antiretroviral agent lopinavir coformulated with ritonavir. *Pharmacogenet Genomics* 2010;**20**:217–30.
109. Zhang X, Tierney C, Albrecht M, et al. Discordant associations between SLCO1B1 521T→C and plasma levels of ritonavir-boosted protease inhibitors in AIDS clinical trials group study A5146. *Ther Drug Monit* 2013;**35**:209–16.
110. Schipani A, Egan D, Dickinson L, et al. Estimation of the effect of SLCO1B1 polymorphisms on lopinavir plasma concentration in HIV-infected adults. *Antivir Ther* 2012; **17**:861–8.
111. Olagunju A, Schipani A, Siccardi M, et al. CYP3A4*22 (c.522-191 C>T; rs35599367) is associated with lopinavir pharmacokinetics in HIV-positive adults. *Pharmacogenet Genomics* 2014;**24**:459–63.
112. Molto J, Xinarianos G, Miranda C, et al. Simultaneous pharmacogenetics-based population pharmacokinetic analysis of darunavir and ritonavir in HIV-infected patients. *Clin Pharmacokinet* 2013;**52**:543–53.
113. Siccardi M, D'Avolio A, Baietto L, et al. Association of a single-nucleotide polymorphism in the pregnane X receptor (PXR 63396C→T) with reduced concentrations of unboosted atazanavir. *Clin Infect Dis* 2008;**47**:1222–5.
114. Schipani A, Siccardi M, D'Avolio A, et al. Population pharmacokinetic modeling of the association between 63396C→T pregnane X receptor polymorphism and unboosted atazanavir clearance. *Antimicrob Agents Chemother* 2010;**54**:5242–50.
115. Albermann N, Schmitz-Winnenthal FH, Z'Graggen K, et al. Expression of the drug transporters MDR1/ABCB1, MRP1/ABCC1, MRP2/ABCC2, BCRP/ABCG2, and PXR in peripheral blood mononuclear cells and their relationship with the expression in intestine and liver. *Biochem Pharmacol* 2005;**70**:949–58.
116. Owen A, Chandler B, Back DJ, Khoo SH. Expression of pregnane-X-receptor transcript in peripheral blood mononuclear cells and correlation with MDR1 mRNA. *Antivir Ther* 2004;**9**:819–21.

Downloaded From Vet-eBooks.com

117. Gammal RS, Court MH, Haidar CE, et al. Clinical pharmacogenetics implementation consortium (CPIC) guideline for UGT1A1 and atazanavir prescribing. *Clin Pharmacol Ther* 2015;**99**(4):363−9.
118. Zucker SD, Qin X, Rouster SD, et al. Mechanism of indinavir-induced hyperbilirubinemia. *Proc Natl Acad Sci USA* 2001;**98**:12671−6.
119. Rotger M, Taffe P, Bleiber G, et al. Gilbert syndrome and the development of antiretroviral therapy-associated hyperbilirubinemia. *J Infect Dis* 2005;**192**:1381−6.
120. Park WB, Choe PG, Song KH, et al. Genetic factors influencing severe atazanavir-associated hyperbilirubinemia in a population with low UDP-glucuronosyltransferase 1A1*28 allele frequency. *Clin Infect Dis* 2010;**51**:101−6.
121. Aceti A, Gianserra L, Lambiase L, Pennica A, Teti E. Pharmacogenetics as a tool to tailor antiretroviral therapy: a review. *World J Virol* 2015;**4**:198−208.
122. Arnedo M, Taffe P, Sahli R, et al. Contribution of 20 single nucleotide polymorphisms of 13 genes to dyslipidemia associated with antiretroviral therapy. *Pharmacogenet Genomics* 2007;**17**:755−64.
123. Tarr PE, Taffe P, Bleiber G, et al. Modeling the influence of APOC3, APOE, and TNF polymorphisms on the risk of antiretroviral therapy-associated lipid disorders. *J Infect Dis* 2005;**191**:1419−26.
124. Egana-Gorrono L, Martinez E, Cormand B, Escriba T, Gatell J, Arnedo M. Impact of genetic factors on dyslipidemia in HIV-infected patients starting antiretroviral therapy. *AIDS* 2013;**27**:529−38.
125. Guardiola M, Ferre R, Salazar J, et al. Protease inhibitor-associated dyslipidemia in HIV-infected patients is strongly influenced by the APOA5-1131T→C gene variation. *Clin Chem* 2006;**52**:1914−9.
126. Li WW, Dammerman MM, Smith JD, Metzger S, Breslow JL, Leff T. Common genetic variation in the promoter of the human apo CIII gene abolishes regulation by insulin and may contribute to hypertriglyceridemia. *J Clin Invest* 1995;**96**:2601−5.
127. Mahley RW, Rall Jr SC. Apolipoprotein E: far more than a lipid transport protein. *Annu Rev Genomics Hum Genet* 2000;**1**:507−37.
128. MacArthur RD, Novak RM. Reviews of anti-infective agents: maraviroc: the first of a new class of antiretroviral agents. *Clin Infect Dis* 2008;**47**:236−41.
129. Lu Y, Fuchs EJ, Hendrix CW, Bumpus NN. CYP3A5 genotype impacts maraviroc concentrations in healthy volunteers. *Drug Metab Dispos* 2014;**42**:1796−802.
130. Vourvahis M, McFadyen L, Heera J, Clark A. Clinical relevance of CYP3A5 genotype on maraviroc exposures. *Drug Metab Dispos* 2015;**43**:771−2.
131. Lu Y, Fuchs EJ, Hendrix CW, Bumpus NN. Response to "clinical relevance of CYP3A5 genotype on maraviroc exposures". *Drug Metab Dispos* 2015;**43**:773.
132. Kassahun K, McIntosh I, Cui D, et al. Metabolism and disposition in humans of raltegravir (MK-0518), an anti-AIDS drug targeting the human immunodeficiency virus 1 integrase enzyme. *Drug Metab Dispos* 2007;**35**:1657−63.
133. Wenning LA, Petry AS, Kost JT, et al. Pharmacokinetics of raltegravir in individuals with UGT1A1 polymorphisms. *Clin Pharmacol Ther* 2009;**85**:623−7.
134. Arab-Alameddine M, Fayet-Mello A, Lubomirov R, et al. Population pharmacokinetic analysis and pharmacogenetics of raltegravir in HIV-positive and healthy individuals. *Antimicrob Agents Chemother* 2012;**56**:2959−66.
135. Calcagno A, Cusato J, Simiele M, et al. High interpatient variability of raltegravir CSF concentrations in HIV-positive patients: a pharmacogenetic analysis. *J Antimicrob Chemother* 2014;**69**:241−5.

Downloaded From Vet-eBooks.com

136. Johnson DH, Sutherland D, Acosta EP, Erdem H, Richardson D, Haas DW. Genetic and non-genetic determinants of raltegravir penetration into cerebrospinal fluid: a single arm pharmacokinetic study. *PLoS One* 2013;**8**:e82672.
137. Moss DM, Siccardi M, Back DJ, Owen A. Predicting intestinal absorption of raltegravir using a population-based ADME simulation. *J Antimicrob Chemother* 2013;**68**:1627−34.
138. Moss DM, Siccardi M, Murphy M, et al. Divalent metals and pH alter raltegravir disposition in vitro. *Antimicrob Agents Chemother* 2012;**56**:3020−6.
139. Chen S, St Jean P, Borland J, et al. Evaluation of the effect of UGT1A1 polymorphisms on dolutegravir pharmacokinetics. *Pharmacogenomics* 2014;**15**:9−16.
140. Navaratnam V, Mansor SM, Sit NW, Grace J, Li Q, Olliaro P. Pharmacokinetics of artemisinin-type compounds. *Clin Pharmacokinet* 2000;**39**:255−70.
141. Ilett KF, Ethell BT, Maggs JL, et al. Glucuronidation of dihydroartemisinin in vivo and by human liver microsomes and expressed UDP-glucuronosyltransferases. *Drug Metab Dispos* 2002;**30**:1005−12.
142. Svensson US, Ashton M. Identification of the human cytochrome P450 enzymes involved in the in vitro metabolism of artemisinin. *Br J Clin Pharmacol* 1999;**48**:528−35.
143. Simonsson US, Lindell M, Raffalli-Mathieu F, Lannerbro A, Honkakoski P, Lang MA. In vivo and mechanistic evidence of nuclear receptor CAR induction by artemisinin. *Eur J Clin Invest* 2006;**36**:647−53.
144. Burk O, Arnold KA, Nussler AK, et al. Antimalarial artemisinin drugs induce cytochrome P450 and MDR1 expression by activation of xenosensors pregnane X receptor and constitutive androstane receptor. *Mol Pharmacol* 2005;**67**:1954−65.
145. Piedade R, Gil JP. The pharmacogenetics of antimalaria artemisinin combination therapy. *Expert Opin Drug Metab Toxicol* 2011;**7**:1185−200.
146. Roederer MW, McLeod H, Juliano JJ. Can pharmacogenomics improve malaria drug policy? *Bull World Health Organ* 2011;**89**:838−45.
147. Kerb R, Fux R, Morike K, et al. Pharmacogenetics of antimalarial drugs: effect on metabolism and transport. *Lancet Infect Dis* 2009;**9**:760−74.
148. Aung AK, Haas DW, Hulgan T, Phillips EJ. Pharmacogenomics of antimicrobial agents. *Pharmacogenomics* 2014;**15**:1903−30.
149. Yusof W, Hua GS. Gene, ethnic and gender influences predisposition of adverse drug reactions to artesunate among Malaysians. *Toxicol Mech Methods* 2012;**22**:184−92.
150. Phompradit P, Muhamad P, Cheoymang A, Na-Bangchang K. Preliminary investigation of the contribution of CYP2A6, CYP2B6, and UGT1A9 polymorphisms on artesunate-mefloquine treatment response in Burmese patients with Plasmodium falciparum malaria. *Am J Trop Med Hyg* 2014;**91**:361−6.
151. Staehli Hodel EM, Csajka C, Ariey F, et al. Effect of single nucleotide polymorphisms in cytochrome P450 isoenzyme and N-acetyltransferase 2 genes on the metabolism of artemisinin-based combination therapies in malaria patients from Cambodia and Tanzania. *Antimicrob Agents Chemother* 2013;**57**:950−8.
152. Zanger UM, Klein K, Saussele T, Blievernicht J, Hofmann MH, Schwab M. Polymorphic CYP2B6: molecular mechanisms and emerging clinical significance. *Pharmacogenomics* 2007;**8**:743−59.
153. Marwa KJ, Schmidt T, Sjogren M, Minzi OM, Kamugisha E, Swedberg G. Cytochrome P450 single nucleotide polymorphisms in an indigenous Tanzanian population: a concern about the metabolism of artemisinin-based combinations. *Malar J* 2014;**13**:420.
154. Maganda BA, Minzi OM, Ngaimisi E, Kamuhabwa AA, Aklillu E. CYP2B6*6 genotype and high efavirenz plasma concentration but not nevirapine are associated with low

Downloaded From Vet-eBooks.com

lumefantrine plasma exposure and poor treatment response in HIV-malaria-coinfected patients. *Pharmacogenomics J* 2016;**16**:88−95.
155. Li XQ, Bjorkman A, Andersson TB, Gustafsson LL, Masimirembwa CM. Identification of human cytochrome P(450)s that metabolise anti-parasitic drugs and predictions of in vivo drug hepatic clearance from in vitro data. *Eur J Clin Pharmacol* 2003;**59**: 429−42.
156. Fletcher KA, Evans DA, Gilles HM, Greaves J, Bunnag D, Harinasuta T. Studies on the pharmacokinetics of primaquine. *Bull World Health Organ* 1981;**59**:407−12.
157. Kim YR, Kuh HJ, Kim MY, et al. Pharmacokinetics of primaquine and carboxyprimaquine in Korean patients with vivax malaria. *Arch Pharm Res* 2004;**27**:576−80.
158. Clayman CB, Arnold J, Hockwald RS, Yount Jr EH, Edgcomb JH, Alving AS. Toxicity of primaquine in Caucasians. *J Am Med Assoc* 1952;**149**:1563−8.
159. Alving AS, Carson PE, Flanagan CL, Ickes CE. Enzymatic deficiency in primaquine-sensitive erythrocytes. *Science* 1956;**124**:484−5.
160. Cappellini MD, Fiorelli G. Glucose-6-phosphate dehydrogenase deficiency. *Lancet* 2008; **371**:64−74.
161. Hill DR, Baird JK, Parise ME, Lewis LS, Ryan ET, Magill AJ. Primaquine: report from CDC expert meeting on malaria chemoprophylaxis I. *Am J Trop Med Hyg* 2006;**75**: 402−15.
162. Li XQ, Bjorkman A, Andersson TB, Ridderstrom M, Masimirembwa CM. Amodiaquine clearance and its metabolism to N-desethylamodiaquine is mediated by CYP2C8: a new high affinity and turnover enzyme-specific probe substrate. *J Pharmacol Exp Ther* 2002; **300**:399−407.
163. White NJ, Looareesuwan S, Edwards G, et al. Pharmacokinetics of intravenous amodiaquine. *Br J Clin Pharmacol* 1987;**23**:127−35.
164. Winstanley P, Edwards G, Orme M, Breckenridge A. The disposition of amodiaquine in man after oral administration. *Br J Clin Pharmacol* 1987;**23**:1−7.
165. Cavaco I, Piedade R, Msellem MI, Bjorkman A, Gil JP. Cytochrome 1A1 and 1B1 gene diversity in the Zanzibar islands. *Trop Med Int Health* 2012;**17**:854−7.
166. Johansson T, Jurva U, Gronberg G, Weidolf L, Masimirembwa C. Novel metabolites of amodiaquine formed by CYP1A1 and CYP1B1: structure elucidation using electrochemistry, mass spectrometry, and NMR. *Drug Metab Dispos* 2009;**37**:571−9.
167. Gil JP, Gil Berglund E. CYP2C8 and antimalaria drug efficacy. *Pharmacogenomics* 2007; **8**:187−98.
168. Parikh S, Ouedraogo JB, Goldstein JA, Rosenthal PJ, Kroetz DL. Amodiaquine metabolism is impaired by common polymorphisms in CYP2C8: implications for malaria treatment in Africa. *Clin Pharmacol Ther* 2007;**82**:197−203.
169. Cavaco I, Martensson A, Froberg G, Msellem M, Bjorkman A, Gil JP. CYP2C8 status of patients with malaria influences selection of Plasmodium falciparum pfmdr1 alleles after amodiaquine-artesunate treatment. *J Infect Dis* 2013;**207**:687−8.
170. Paganotti GM, Gallo BC, Verra F, et al. Human genetic variation is associated with Plasmodium falciparum drug resistance. *J Infect Dis* 2011;**204**:1772−8.
171. Fontaine F, de Sousa G, Burcham PC, Duchene P, Rahmani R. Role of cytochrome P450 3A in the metabolism of mefloquine in human and animal hepatocytes. *Life Sci* 2000;**66**: 2193−212.
172. Pham YT, Regina A, Farinotti R, et al. Interactions of racemic mefloquine and its enantiomers with P-glycoprotein in an immortalised rat brain capillary endothelial cell line, GPNT. *Biochim Biophys Acta* 2000;**1524**:212−9.

Downloaded From Vet-eBooks.com

173. Aarnoudse AL, van Schaik RH, Dieleman J, et al. MDR1 gene polymorphisms are associated with neuropsychiatric adverse effects of mefloquine. *Clin Pharmacol Ther* 2006;**80**:367−74.
174. Kerb R. Implications of genetic polymorphisms in drug transporters for pharmacotherapy. *Cancer Lett* 2006;**234**:4−33.
175. Piedade R, Traub S, Bitter A, et al. Carboxymefloquine, the major metabolite of the antimalarial drug mefloquine, induces drug-metabolizing enzyme and transporter expression by activation of pregnane X receptor. *Antimicrob Agents Chemother* 2015;**59**: 96−104.
176. Herrlin K, Massele AY, Rimoy G, et al. Slow chloroguanide metabolism in Tanzanians compared with white subjects and Asian subjects confirms a decreased CYP2C19 activity in relation to genotype. *Clin Pharmacol Ther* 2000;**68**:189−98.
177. Kaneko A, Kaneko O, Taleo G, Bjorkman A, Kobayakawa T. High frequencies of CYP2C19 mutations and poor metabolism of proguanil in Vanuatu. *Lancet* 1997;**349**: 921−2.
178. Janha RE, Sisay-Joof F, Hamid-Adiamoh M, et al. Effects of genetic variation at the CYP2C19/CYP2C9 locus on pharmacokinetics of chlorcycloguanil in adult Gambians. *Pharmacogenomics* 2009;**10**:1423−31.
179. Edstein MD, Yeo AE, Kyle DE, Looareesuwan S, Wilairatana P, Rieckmann KH. Proguanil polymorphism does not affect the antimalarial activity of proguanil combined with atovaquone in vitro. *Trans R Soc Trop Med Hyg* 1996;**90**:418−21.
180. Kaneko A, Bergqvist Y, Taleo G, Kobayakawa T, Ishizaki T, Bjorkman A. Proguanil disposition and toxicity in malaria patients from Vanuatu with high frequencies of CYP2C19 mutations. *Pharmacogenetics* 1999;**9**:317−26.
181. Song T, Chen J, Huang L, et al. Should we abandon quinine plus antibiotic for treating uncomplicated falciparum malaria? A systematic review and meta-analysis of randomized controlled trials. *Parasitol Res* 2015;**115**(3):903−12.
182. *Guidelines for the treatment of malaria*. 3rd ed. 2015.
183. Taylor WR, White NJ. Antimalarial drug toxicity: a review. *Drug Saf* 2004;**27**:25−61.
184. Zhang H, Coville PF, Walker RJ, Miners JO, Birkett DJ, Wanwimolruk S. Evidence for involvement of human CYP3A in the 3-hydroxylation of quinine. *Br J Clin Pharmacol* 1997;**43**:245−52.
185. Mirghani RA, Sayi J, Aklillu E, et al. CYP3A5 genotype has significant effect on quinine 3-hydroxylation in Tanzanians, who have lower total CYP3A activity than a Swedish population. *Pharmacogenet Genomics* 2006;**16**:637−45.
186. Dooley KE, Phillips PP, Nahid P, Hoelscher M. Challenges in the clinical assessment of novel tuberculosis drugs. *Adv Drug Deliv Rev* 2016;**102**:116−22.
187. Hughes HB. On the metabolic fate of isoniazid. *J Pharmacol Exp Ther* 1953;**109**:444−52.
188. Hughes HB, Biehl JP, Jones AP, Schmidt LH. Metabolism of isoniazid in man as related to the occurrence of peripheral neuritis. *Am Rev Tuberc* 1954;**70**:266−73.
189. Ramachandran G, Swaminathan S. Role of pharmacogenomics in the treatment of tuberculosis: a review. *Pharmgenomics Pers Med* 2012;**5**:89−98.
190. Matsumoto T, Ohno M, Azuma J. Future of pharmacogenetics-based therapy for tuberculosis. *Pharmacogenomics* 2014;**15**:601−7.
191. Cai Y, Yi J, Zhou C, Shen X. Pharmacogenetic study of drug-metabolising enzyme polymorphisms on the risk of anti-tuberculosis drug-induced liver injury: a meta-analysis. *PLoS One* 2012;**7**:e47769.

Downloaded From Vet-eBooks.com

192. Wang PY, Xie SY, Hao Q, Zhang C, Jiang BF. NAT2 polymorphisms and susceptibility to anti-tuberculosis drug-induced liver injury: a meta-analysis. *Int J Tuberc Lung Dis* 2012; **16**:589−95.
193. Donald PR, Sirgel FA, Venter A, et al. The influence of human N-acetyltransferase genotype on the early bactericidal activity of isoniazid. *Clin Infect Dis* 2004;**39**:1425−30.
194. Jung JA, Kim TE, Lee H, et al. A proposal for an individualized pharmacogenetic-guided isoniazid dosage regimen for patients with tuberculosis. *Drug Des Devel Ther* 2015;**9**: 5433−8.
195. Du H, Chen X, Fang Y, et al. Slow N-acetyltransferase 2 genotype contributes to anti-tuberculosis drug-induced hepatotoxicity: a meta-analysis. *Mol Biol Rep* 2013;**40**:3591−6.
196. Sun F, Chen Y, Xiang Y, Zhan S. Drug-metabolising enzyme polymorphisms and predisposition to anti-tuberculosis drug-induced liver injury: a meta-analysis. *Int J Tuberc Lung Dis* 2008;**12**:994−1002.
197. Leiro-Fernandez V, Valverde D, Vazquez-Gallardo R, et al. N-acetyltransferase 2 polymorphisms and risk of anti-tuberculosis drug-induced hepatotoxicity in Caucasians. *Int J Tuberc Lung Dis* 2011;**15**:1403−8.
198. Leiro V, Fernandez-Villar A, Valverde D, et al. Influence of glutathione S-transferase M1 and T1 homozygous null mutations on the risk of antituberculosis drug-induced hepatotoxicity in a Caucasian population. *Liver Int* 2008;**28**:835−9.
199. Chatterjee S, Lyle N, Mandal A, Kundu S. GSTT1 and GSTM1 gene deletions are not associated with hepatotoxicity caused by antitubercular drugs. *J Clin Pharm Ther* 2010;**35**: 465−70.
200. Kim SH, Kim SH, Yoon HJ, et al. GSTT1 and GSTM1 null mutations and adverse reactions induced by antituberculosis drugs in Koreans. *Tuberc (Edinb)* 2010;**90**:39−43.
201. Weiner M, Peloquin C, Burman W, et al. Effects of tuberculosis, race, and human gene SLCO1B1 polymorphisms on rifampin concentrations. *Antimicrob Agents Chemother* 2010;**54**:4192−200.
202. Chigutsa E, Visser ME, Swart EC, et al. The SLCO1B1 rs4149032 polymorphism is highly prevalent in South Africans and is associated with reduced rifampin concentrations: dosing implications. *Antimicrob Agents Chemother* 2011;**55**:4122−7.
203. Hennig S, Naiker S, Reddy T, et al. Effect of SLCO1B1 polymorphisms on rifabutin pharmacokinetics in African HIV-infected patients with tuberculosis. *Antimicrob Agents Chemother* 2015;**60**:617−20.
204. McIlleron H, Abdel-Rahman S, Dave JA, Blockman M, Owen A. Special populations and pharmacogenetic issues in tuberculosis drug development and clinical research. *J Infect Dis* 2015;**211**(Suppl. 3):S115−25.

Downloaded From Vet-eBooks.com

This page intentionally left blank

Downloaded From Vet-eBooks.com

Genetic Exchange in Trypanosomatids and Its Relevance to Epidemiology

20

W. Gibson [1], *M.D. Lewis* [2], *M. Yeo* [2], *M.A. Miles* [2]
[1]University of Bristol, Bristol, United Kingdom; [2]London School of Hygiene and Tropical Medicine (LSHTM), London, United Kingdom

1. Introduction

Genetic exchange has now been demonstrated experimentally in the so-called TriTryps, three of the trypanosomatids for which genome sequences have been published[1–3]: *Trypanosoma brucei*, *Trypanosoma cruzi*, *Leishmania major*, and also within and between other *Leishmania* species.[4] As well as in these human pathogens, the related bumble bee parasite, *Crithidia bombi*, has also been shown to undergo genetic exchange.[5] For *T. brucei* and *Leishmania* spp. genetic exchange occurs in the insect vector and appears to follow Mendelian rules of inheritance.[6–8] In contrast, for *T. cruzi* genetic exchange has been demonstrated experimentally in infected mammalian cell cultures, and hybrid formation appears to result from fusion of diploid cells followed by genome erosion.[9]

Before the experimental confirmation of genetic exchange in these parasites, evidence for the natural occurrence of hybrids had accumulated from molecular epidemiological analysis of isolates collected from the field. Tait[10] showed that isoenzyme data from *T. brucei* conformed to Hardy–Weinberg equilibrium and concluded that the population was undergoing random mating. Later analyses threw doubt on the extent of panmictic mating in *T. brucei* with the widespread acceptance of the concept of clonality in parasitic protozoa.[11] However, genetic exchange was considered to have led to demonstrably hybrid *T. cruzi* recovered from the field,[12–15] including TcV and TcVI interlineage natural hybrids, subsequently indicated to be of relatively recent origin.[16] Natural hybrids were also demonstrated between several *Leishmania* species and subspecies.[17–19]

Here, we review results from these two complementary avenues of study—analysis of naturally occurring hybrids among field isolates and experimental genetic exchange in the laboratory—for the pathogenic trypanosomatids, *T. brucei*, *T. cruzi*, and *L. major*.

2. *Trypanosoma brucei*

2.1 Genetic Crosses

Compelling evidence for mating in *T. brucei sensu lato* came from population genetics studies based on isoenzyme data. Gibson et al.[20] described isoenzyme patterns

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00020-2
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

consistent with those expected from homo- and heterozygotes in an extensive analysis of *T. brucei* isolates from East and West Africa, and Tait[10] showed that similar data from 17 Ugandan isolates of *T. brucei* conformed to Hardy−Weinberg equilibrium, indicating that the population was undergoing random mating. The first successful laboratory cross was reported in 1986 when Jenni et al. cotransmitted two genetically distinct clones of *T. brucei s.l.* through tsetse flies and demonstrated hybrid progeny that had inherited a mixture of genetic markers from both parents.[7] The hybrids were found among metacyclics recovered from the salivary glands showing that genetic exchange had occurred sometime during the trypanosome's developmental cycle in the tsetse fly. However, the metacyclic population contained a mixture of parental and hybrid genotypes indicating that mating is not an obligatory event in the life cycle. This contrasts with the situation for the malaria parasite, *Plasmodium* spp., where gamete formation and production of a zygote is a normal part of the transmission cycle.

Subsequent trypanosome crosses have followed the same general plan: two genetically distinct parental trypanosome clones are fed to groups of newly emerged tsetse flies in their first blood meal (Fig. 20.1). Tsetse flies are typically refractory to trypanosome infection, but are at their most susceptible as very young flies before they have fed.[21] Not all flies become infected after the infected feed and only some infected flies produce hybrids. Thus large numbers of flies and trypanosome populations need to be screened to identify those containing hybrids. To avoid the laborious job of identifying hybrid-producing flies by screening every fly, selectable markers were incorporated into the experimental design. This became feasible following the development of methods for the stable transformation of trypanosomes

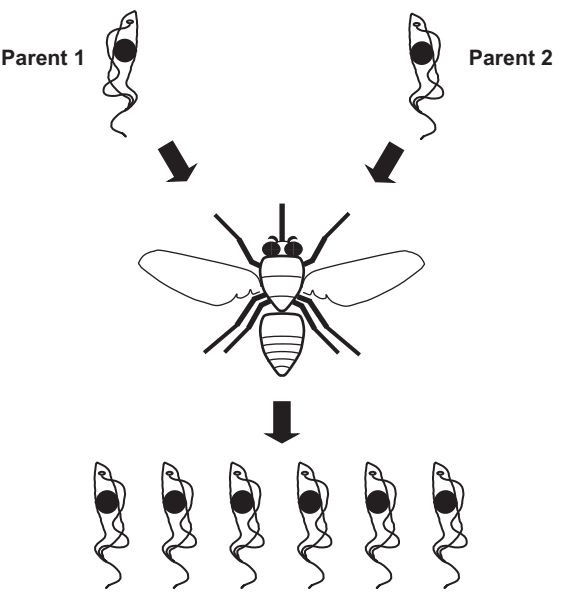


Figure 20.1 Design of an experimental cross for *Trypanosoma brucei*. The two parental trypanosomes are cotransmitted via tsetse flies, and hybrid trypanosomes are found among the infective metacyclics from the salivary glands.

Downloaded From Vet-eBooks.com

with exogenous DNA in the early 1990s.[22–24] In the cross described by Gibson and Whittington,[25] each of the parental clones was transformed with a different construct designed to integrate a gene for drug resistance into the tubulin locus by homologous recombination. In this way, parental clones resistant to the antibiotics hygromycin or G418 were created. After cotransmission through the fly, hybrid progeny were selected by resistance to both drugs. This strategy has obvious advantages over the previous "finding a needle in a haystack" approach, and was used for the discovery of hybrids in *L. major*.[6]

The discovery of green fluorescent protein (GFP) and the development of methods to image the protein in living cells opened up exciting new approaches for studying genetic exchange in trypanosomes. Incorporation of genes for GFP in one parental line and red fluorescent protein (RFP) in the other led to the production of hybrids with both genes, which appear yellow making them immediately distinct from the parental cells by simple fluorescence microscopy of live cell preparations. This approach was a boon for the analysis of *T. brucei* hybrids within the tsetse fly and allowed the location and timing of genetic exchange to be determined precisely.[26]

It has proved possible to cross all subspecies of *T. brucei* except the human pathogen *T. brucei gambiense* Group 1. The difficulty in setting up crosses of *T. b. gambiense* Group 1 is that it transmits poorly or not at all through *Glossina morsitans* ssp., the standard laboratory tsetse fly.[27,28] The more virulent Group 2 *T. b. gambiense* is readily transmissible via *G. morsitans* ssp. and has featured in several crosses, including the original cross of Jenni et al.,[7] where the parents were *T. brucei brucei* STIB 247 and *T. b. gambiense* Group 2 STIB 386 (TH114). *T. b. gambiense* Group 2 (TH2) was also mated with both *T. b. brucei* and *T. brucei rhodesiense*.[29,30] Several crosses have involved parental lines of *T. b. brucei*,[26,31–34] or *T. b. brucei* and *T. b. rhodesiense*.[25,35,36] Inheritance of the trait for human infectivity has been analyzed in crosses of *T. b. brucei* and *T. b. rhodesiense*,[37] and *T. b. brucei* and Group 2 *T. b. gambiense*.[38] A single gene, *SRA* (serum resistance associated), confers human infectivity on *T. b. rhodesiense*,[39,40] and it was possible to follow the inheritance of this gene, along with the trait of human serum resistance, in crosses of *T. b. brucei* and *T. b. rhodesiense*.[35]

2.2 Location of Genetic Exchange

Jenni et al.[7] originally showed that genetic exchange took place in the fly, since cloned metacyclics from the saliva of infected flies were hybrid; however, the exact location and lifecycle stage remained undefined. While there were reports of hybrid formation in mixed midgut procyclic populations both in vitro and in vivo,[41–43] the bulk of evidence pointed to the salivary glands as the site of genetic exchange. Firstly, the timing of hybrid appearance: hybrids were most likely to be found in flies infected for at least 28 days, that is, sufficient time for salivary gland invasion and colonization, despite there being a large population of procyclic trypanosomes continuously present in the midgut throughout this time.[25,30,44] Selection by double-drug resistance revealed that hybrids were present in populations derived from the salivary glands but not from the midgut.[25,30,45] The direct visualization of trypanosome hybrids using fluorescent

Downloaded From Vet-eBooks.com

reporter proteins unequivocally established that hybrids are formed in the salivary glands and are not present in the midgut or among the migratory stages in the proventriculus and foregut.[26,29]

The use of red and green fluorescent reporters also explained why hybrids are found infrequently, with less than a quarter of infected flies producing hybrids.[46] In order to mate, both parental trypanosomes need to reach and colonize the same salivary gland, and it became evident that this is not always the case. While most flies coinfected with red and green fluorescent trypanosomes developed a mixed midgut infection, only about a third of these flies also had a mixed infection in the salivary glands.[26,47] In a number of cases, the composition of the trypanosome population in the two salivary glands of the pair differed, perhaps with only one salivary gland infected, or a mixed infection in one gland but not the other. Interestingly, all progeny from the red/green cross were hybrid and no trypanosomes with parental genotypes were recovered as in the previous crosses.[26] The design of this cross may have increased the probability of finding hybrids, as analysis could be focused on salivary glands containing both parental genotypes, not possible in previous crosses.

These observations highlight the fact that few trypanosomes complete the journey from the midgut to the salivary duct, and only some of these then succeed in establishing an infection in either of the salivary glands. When together as epimastigotes in the same salivary gland, the trypanosomes readily mate, as demonstrated by the fact that most salivary glands with a mixed infection of red and green trypanosomes also contained yellow fluorescent hybrids.[26] Compatibility of different trypanosome strains may then depend on them reaching the salivary glands simultaneously. In some early crosses, the two parental clones differed substantially in their speed of colonizing the salivary glands.[25,30,44] The fact that mating in *T. brucei* occurs among epimastigotes in the salivary glands makes the prospect of producing hybrids in vitro more remote, as reliable culture systems do not exist for the life cycle stages that occur in the salivary glands.

2.3 Mendelian Inheritance and Meiosis

It is generally accepted that *T. brucei* is diploid with respect to the 11 pairs of large chromosomes that contain the housekeeping genes, although this arrangement probably does not apply to the intermediate and minichromosomes.[1,48,49] Analysis of the inheritance of genetic markers in the hybrid progeny from crosses of *T. brucei* is consistent with Mendelian genetics for the most part,[8,33,34,36,45,50] leading to the assumption that meiosis occurs. Indeed, genetic linkage maps for *T. b. brucei* and Group 2 *T. b. gambiense* have been constructed from detailed analysis of microsatellite inheritance and frequency of crossing over.[51,52]

It has proved more difficult to directly visualize trypanosomes undergoing meiosis. While hybrids were easily detected in a cross of red and green fluorescent trypanosomes from day 13 onward,[26] putative intermediate stages were neither abundant nor obvious, necessitating an alternative approach to detect trypanosomes undergoing meiosis. Phylogenomic studies had identified the presence, in *T. brucei*, of homologues of several genes crucial for meiosis in yeast and in other eukaryotes, such as

Downloaded From Vet-eBooks.com

Spo11, Hop1, Dmc1, and Mnd1.[53] By constructing fusions of the *T. brucei* homologues with the gene for yellow fluorescent protein (YFP), expression of these putative meiosis genes was monitored through the developmental cycle of *T. brucei* in the fly. Expression was observed only among trypanosomes during the early stage of colonization of the salivary glands, consistent with the first appearance of hybrids and was localized to the nucleus.[54] Surprisingly, the meiotic trypanosomes occurred with similar frequency in both single and mixed strain transmissions and it seems probable that meiosis is a normal part of the trypanosome developmental cycle in the fly. Expression of meiosis-specific genes has now been observed in all subspecies of *T. brucei*, including *T. b. gambiense* Group 1.[55] The identification of trypanosomes undergoing meiosis was the first step in elucidating the mechanism of genetic exchange in *T. brucei* at the cell biology level, and eventually led to the identification of haploid gametes[55]; this stage had initially proved elusive.[56–58] Observation of trypanosomes from the early phase of salivary gland colonization, when meiotic stages are present, revealed the presence of distinctive pear-shaped cells with a long flagellum; in mixtures of red and green fluorescent trypanosomes, these cells were seen to interact by intertwining their flagella, and transfer of cytoplasmic material was also evident as some cells had yellow fluorescence.[55] Measurement of DNA contents of individual nuclei showed that the pear-shaped cells had half the nuclear DNA content of metacyclics, which are nondividing trypanosomes arrested in G_0. Curiously, about half the haploid cells observed had one nucleus and two kinetoplasts (1N2K), but only a single flagellum; the rest were 1N1K cells with one flagellum.[55] It is not yet clear how these two types of haploid gametes arise from the trypanosomes undergoing meiosis and what intermediates are involved. The meiosis-specific genes that trypanosomes express (*MND1*, *DMC1*, and *HOP1*) are characteristic of meiosis I, and therefore another round of division should follow, yielding four haploid nuclei. Whether this division also involves replication of the kinetoplast, basal body, and flagellum is not known.

Previously, indirect evidence for haploid nuclei came from the observation that many *T. brucei* ssp. crosses produced triploid hybrid progeny, which most probably arose from fusion of a haploid nucleus with one that is diploid.[25,30,45,59,60] Hybrids with high DNA contents relative to the parents were found even in the first experimental cross,[7] which created some initial confusion about the mechanism of genetic exchange.[61,62] Analysis of progeny clones with high DNA contents from several further crosses demonstrated that these hybrids were triploid with DNA contents that clustered at the 3N value, and, in addition, trisomy was confirmed for several chromosomes.[25,30,45,59,60]

As well as triploid hybrids, several tetraploid hybrids were recovered from a cross of red and green fluorescent trypanosomes.[26] While the presence of triploid hybrids was obvious from the demonstration of three alleles at some loci, only two microsatellite alleles were present in each of the tetraploid hybrids. They were therefore not formed by the fusion of the two diploid parental genomes, but appear to be the products of genome endoreplication, for example, by fusion of gametes before they have undergone reduction division at the end of meiosis I.[26] While triploids appear to be stable during growth and fly transmission,[59,62] the tetraploids may be unstable, as

Downloaded From Vet-eBooks.com

flow cytometry analysis of the DNA contents of tetraploid clones frequently revealed an extra G1 peak at the 2N position.[26]

Intraclonal mating was initially thought not to occur in *T. brucei* except in the presence of outcrossing trypanosomes, leading to the hypothesis that some kind of diffusible factor produced by non-self-recognition induced mating.[30,63] Intraclonal mating explained the occasional anomalies where hybrid progeny were homozygous instead of heterozygous as expected if the parents were different homozygotes.[45,64] Experiments reported in 2009 using red and green fluorescent clones of a single *T. brucei* strain have shown that intraclonal mating occurs with some frequency in the absence of a second trypanosome strain.[65] The assumption that a *T. brucei* clone can be tsetse-transmitted without change is therefore doubtful.

In other protists, such as ciliates, strain compatibility is determined by a system of mating types, but the question of trypanosome mating types remains open. Analysis of a series of F1 and back crosses revealed no systematic pattern of compatible matings.[66] When red and green haploid gametes from intraclonal crosses were observed, they behaved exactly like the gametes from interclonal crosses, coming into close proximity and intertwining their flagella, except that no mixing of cytoplasm (yellow fluorescence) was evident among gametes of the same trypanosome strain.[66]

2.4 Inheritance of Kinetoplast DNA

Kinetoplast DNA (kDNA) is the mitochondrial DNA of trypanosomatids and consists of an interlocked network of about 50 (20−25 kb) maxicircles and 5000 (1 kb) minicircles (reviewed by [67]). The kDNA is contained within an organelle, the kinetoplast, which is inside the mitochondrion. Initial results supported the hypothesis that inheritance of kDNA was uniparental, with the kDNA of either parent being passed on to the hybrid progeny.[36,50,64] However, detailed analysis of both maxi- and minicircles showed that although maxicircles were of a single parental type, the minicircles had been inherited from both parents.[68,69] To explain this result, it was assumed that the hybrid kDNA network initially consists of both maxi- and minicircles from both parents in an equal proportion; the small number of maxicircles resolves to a single type by random segregation during subsequent mitotic divisions, while the much greater number of minicircles endures as a hybrid network. This idea is supported by the observation that hybrids at an early stage of growth have mixed maxicircle networks.[26,70]

The initial formation of the hybrid kDNA network remains an intriguing problem. The first requirement is fusion of the mitochondria from both parental trypanosomes to allow the kDNA to mix. Then mini- and maxicircles from both parental kDNA networks would need to combine into a single hybrid kDNA network, which would involve detachment and reattachment of all the individual DNA circles. An alternative hypothesis is that only some minicircles are swapped while the kDNA networks of the two parents are adjacent, leaving the core maxicircle network intact; the partially hybrid kDNA networks would then be inherited by individual progeny trypanosomes.[67] However, this hypothesis does not fit with the observation of mixed maxicircle networks mentioned earlier.

Downloaded From Vet-eBooks.com

2.5 Implications for Epidemiology

Although genetic exchange in *T. brucei* has been amply demonstrated in the laboratory, its importance in natural trypanosome populations remains controversial. One problem has been sampling bias, with collection of trypanosome isolates usually focused on epidemics of human disease rather than collecting all the *T. brucei* strains circulating in an area.[71–74] In epidemics of human trypanosomiasis caused by *T. b. rhodesiense*, transmission may well be direct from human to human rather than from an animal reservoir, or involve only local cattle reservoir hosts, allowing the clonal expansion of particular trypanosome genotypes. This is very different from the endemic scenario where humans and their livestock are occasional hosts in the natural circulation of *T. brucei* ssp. strains in wild mammals and tsetse.

Through the analysis of trypanosome mating in the laboratory, we can now define the biological circumstances in which genetic exchange will be found: at least two trypanosome strains must be present in the salivary glands of a tsetse fly for mating to occur; since flies are most readily infected on their first blood meal,[21,75] a mixed infection is likely to be acquired from one infected mammal carrying multiple trypanosome strains. Few mixed infections have been reported from humans,[76,77] but mixed infections of more than one trypanosome strain or species are frequently encountered in livestock,[78,79] tsetse,[80–83] and presumably also occur frequently in the large wild mammals that sustain many tsetse populations. In-depth analysis of trypanosome samples from these transmission cycles would be informative.

Even though the frequency of genetic exchange in natural populations may be low, there is potential for significant epidemiological consequences. For example, the trait of human infectivity in *T. b. rhodesiense* is conferred by a single gene, *SRA*.[39,40] Any cross between *T. b. rhodesiense* and *T. b. brucei* would place this key virulence gene in new genetic backgrounds, thus creating new genotypes of *T. b. rhodesiense*. There is abundant evidence of strain heterogeneity in *T. b. rhodesiense* from several foci of human trypanosomiasis in East Africa,[73,84–87] and quite different *T. b. rhodesiense* genotypes have been found in neighboring foci in Uganda and Kenya.[85,86,88] We know little about the genetic basis of most phenotypic characteristics of *T. brucei s.l.*, but undoubtedly, genetic exchange provides the opportunity for more virulent, pathogenic, or fly-transmissible strains of *T. b. rhodesiense* to arise. Population genetics analysis of *T. brucei* ssp. strains from across Africa provided strong evidence of genetic admixture between *SRA*-positive *T. b. rhodesiense* and *SRA*-negative *T. b. brucei* in East Africa,[89] and transfer of the *SRA* gene was demonstrated in several *T. b. rhodesiense* x *T. b. brucei* laboratory crosses, thus creating new genetic strains of the human pathogen.[35]

3. *Trypanosoma cruzi*

3.1 Trypanosoma cruzi *Diversity*

Trypanosoma cruzi is considered to be a single species, but is comprised of six distinct genetic lineages (also referred to as discrete typing units, DTUs). The definition of

Downloaded From Vet-eBooks.com

these six *T. cruzi* subgroups was originally on the basis of phenotyping using multilocus enzyme electrophoresis (MLEE).[90,91] Subsequently, the same six genetic lineages were supported by comparative analyses of a wide range of nuclear DNA targets. Messenger et al.[92] provide a history of research on the diversity of *T. cruzi*, and the methods used in analyzing such diversity have also been described in detail.[91] The current international consensus nomenclature for the six *T. cruzi* lineages is TcI–TcVI.[93]

TcI–TcVI show broadly distinctive but not entirely exclusive geographical, ecological, transmission cycle and disease associations; overlaps occur, and mixed infections are reported from humans, reservoir mammal hosts, and arthropod vectors.[92,94–98] TcI is the principal agent of Chagas disease in Latin America north of the Amazon basin, whereas TcII, TcV, and TcVI are the main causes of Chagas disease in the Southern Cone countries of South America.[92,99–103] The sylvatic TcI transmission cycles are widespread throughout Latin America and are largely arboreal; the common opossums (*Didelphis* species) are obvious and abundant reservoir hosts but many other mammal species may be infected, with some transmission among rodent species and triatomines with terrestrial habitats.[97] The natural transmission cycles of TcII, TcV, and TcVI appear to be rare or difficult to discover and characterize; however, the presence of TcII in primates of the Atlantic forest region of Brazil has been confirmed by a combination of genotyping and *T. cruzi* lineage–specific serology[104,105] (Kerr et al., unpublished). TcIII is predominantly sylvatic and seldom infects humans. TcIV is a secondary cause of Chagas disease in Venezuela, after TcI.[106] TcIII appears to have the most clearly defined and exclusive natural ecological and host association, with the burrowing armadillo, *Dasypus novemcinctus*, and rarely infects humans.[92,97]

The various *T. cruzi* isolates that were characterized to establish this intraspecific genetic diversity were collected from disparate geographical locations and usually with small numbers of isolates gathered from each site. Although this striking genetic heterogeneity of *T. cruzi* was a landmark discovery, which has transformed subsequent research on the epidemiology of Chagas disease, this sampling strategy was not ideally suited for determining the structure of *T. cruzi* populations. Nevertheless, repeated analyses of the molecular diversity of *T. cruzi* based on such samples have supported the discrete nature of the six genetic lineages and demonstrated strong linkage disequilibrium between them.[11] These observations fostered the tenet that *T. cruzi* was propagated clonally, both between and within the genetic lineages, and that genetic exchange, if occurred, was rare and of little epidemiological consequence.[107] On the other hand, despite this clonal theory, even the early MLEE analyses indicated that TcV and TcVI resembled natural interlineage hybrids between TcII and TcIII.[108] The application of multilocus DNA sequencing eventually confirmed that TcV and TcVI are hybrid lineages,[13,109] and probably with a relatively recent origin.[16] However, this was not before the TcVI reference strain "CL Brener" had been selected for the first attempt at obtaining a full *T. cruzi*–genome sequence, as part of the TriTryp genome sequencing project. As discussed later, TcV and TcVI are endemic agents of Chagas disease across much of the Southern Cone region, making the study of recombination in *T. cruzi* of profound epidemiological relevance.

Downloaded From Vet-eBooks.com

3.2 Genome Sequence of a Natural Hybrid

The hybrid nature of the CL Brener strain made the task of sequencing its genome particularly challenging since most genes were represented by two divergent copies.[2] This complicated the assembly of the genome and eventually required additional sequence data from a representative of one of the parental groups, for which the TcII strain "Esmeraldo" was chosen. This allowed the putative parental TcII ("Esmeraldo-like") and TcIII ("non-Esmeraldo-like") sequences to be partially deduced from the single hybrid genome sequence. The initial assembly of the genome was further complicated by the large number of repetitive surface protein gene families that were found throughout the *T. cruzi* genome. The CL Brener haploid genome contained about 12,000 genes and was about 55 Mb in size, considerably larger than either *T. brucei* (26 Mb),[1] or *L. major* (33 Mb).[3] The reassembled CL Brener sequence has produced fewer, larger contigs compared to the original assembly.[110] Other *T. cruzi* genomes have now been sequenced and focus on the remaining genetic lineages, which have less complex genomes.[111,112] Due to its degree of divergence and epidemiological importance as an agent of Chagas disease, particularly north of the Amazon, TcI is of special interest. With the aid of new sequencing technologies, the TcI genome has now been assembled to virtually chromosome-sized contigs, and with high resolution of complex multiple gene families (Talavares et al., in preparation).

3.3 Genetic Crosses

In contrast with *T. brucei*, only a single successful genetic crossing experiment has so far been reported for *T. cruzi*. The earliest, and unsuccessful, attempts at genetic crosses involved passaging the TcI and TcII together in vitro, through triatomine bugs, or through mice, followed by phenotypic analysis of the resultant populations.[108] As with *T. brucei*, such experiments were revolutionized by the ability to genetically transform *T. cruzi* strains to carry different drug resistance markers, permitting selection of any double drug−resistant populations emerging from genetic crossing experiments. Perhaps the best experimental strategy for application of this new technology would be an attempt to cross *T. cruzi* strains from within the same DTU, an approach encouraged by the discovery of both putative hybrid and parental TcI phosphoglucomutase isoenzyme phenotypes among clones of *T. cruzi* isolates from a single undisturbed locality in Amazonian forest.[12] Accordingly, a pair of these putative TcI parental isolates was selected, biological clones were prepared, and they were transfected to carry different episomal drug resistance markers for hygromycin and neomycin (G418). The transgenic parental isolates were passaged together through triatomine bugs, mice, and mammalian cell cultures, and the recovered populations were cultured in a media containing both the drugs to select possible hybrids. The only double drug−resistant clones were derived from the mammalian cell cultures.[113] MLEE, karyotyping, gene sequencing, and microsatellite profile analysis demonstrated that the six double drug−resistant clones were hybrids of the two parental strains.[9] Thus, it was demonstrated experimentally for the first time that *T. cruzi* has an extant capacity to undergo genetic exchange, at least within TcI.

Downloaded From Vet-eBooks.com

The availability of multiple fluorescent protein markers with distinct excitation and emission spectra provides a means to study genetic exchange in greater detail. As described earlier, crosses between *T. brucei* cell lines carrying either a GFP or an RFP gene led to readily identifiable hybrid organisms expressing both reporters such that they appear yellow in composite images of tsetse salivary glands.[26] Red and green fluorescent lines of *T. cruzi* and the closely related species *T. rangeli*, with strong and stable expression of fluorescence and drug resistance markers, have been generated by integrating reporter constructs into either the ribosomal RNA gene array[114] or the tubulin locus.[115] Just as *T. brucei* mating only occurs between strains residing in the same tsetse fly salivary gland, it is expected that experimental crosses in *T. cruzi* will only be successful if two strains can be brought into close physical proximity at the appropriate point of their life cycles. The use of red and green fluorescently labeled *T. cruzi* cell lines has already enabled the tracking of coinfections of mammalian cell cultures in vitro and in mice and triatomine bugs in vivo indicating that there should be few technical barriers in identifying coexpressing "yellow" hybrids[114,115] (Fig. 20.2). Indeed, mammalian cell cultures simultaneously infected with several different pairs of red and green transgenic strains commonly display coinfections of individual cells.

3.4 Location of Genetic Exchange

In the successful experimental *T. cruzi* cross described by Gaunt et al.,[9] nonconfluent mammalian cell cultures were infected with *T. cruzi* cultures containing both infective metacyclic trypomastigotes and noninfective epimastigotes. Once the metacyclic trypomastigotes had invaded the mammalian cells, the epimastigotes were removed by washing the cell monolayers. The infected mammalian cell cultures were maintained for more than 20 days, allowing completion of several rounds of intracellular replication, each of which may take as little as 5 days. The double drug−resistant hybrids were recovered from the culture supernatant, which contained trypomastigotes released from lysed mammalian cells. Thus, the most obvious interpretation is that hybridization took place intracellularly in a coinfected host cell, and that it therefore involved either amastigotes or trypomastigotes. However, it is also conceivable that genetic exchange could have taken place between extracellular forms prior to host cell invasion (i.e., epimastigotes or metacyclic trypomastigotes), or after infected cells had ruptured (i.e., new trypomastigotes, including short or slender forms).

In terms of opportunity in endemic areas, where natural recombinants do exist, mixed interlineage *T. cruzi* infections and multiclonal intralineage infections are by no means rare in humans or other mammals.[9,92,94,116−118] Furthermore, sylvatic mammals are likely to be subject to multiple challenge infections, some of which may be orally acquired by consuming triatomine bugs. *Didelphis* can have anal gland infections that include morphological stages typically restricted to the insect vector. Nevertheless, by analogy with *T. brucei* and *Leishmania* where genetic exchange takes place in the tsetse or sand fly vectors, respectively, it would be surprising if genetic exchange in *T. cruzi* did not occur in its insect vector. There would be abundant opportunity for *T. cruzi* to undergo genetic exchange in triatomine bugs. While tsetse

Downloaded From Vet-eBooks.com

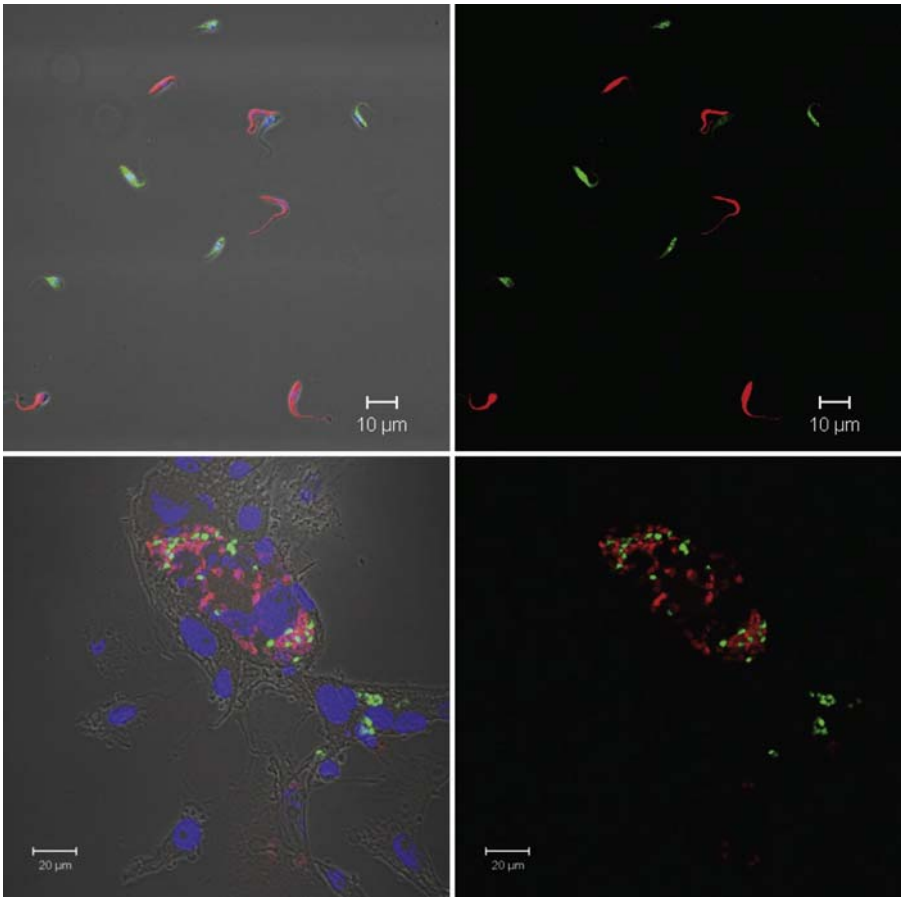


Figure 20.2 Transgenic *Trypanosoma cruzi* cell lines expressing red or green fluorescent protein markers. Top panels show mixed samples of extracellular flagellate forms; bottom panels show mammalian cells coinfected with amastigote forms. Panels on the left show phase microscopy, with red fluorescence, green fluorescence, and DNA staining; panels on the right show red and green fluorescence only. DNA is stained with Hoechst 33342, a blue fluorescent dye.

flies have limited susceptibility to *T. brucei* and most readily acquire infection during their first feed, triatomine bugs may acquire *T. cruzi* infection at any of the five nymphal stages or as adults. Each bug takes many feeds, any one of which may be infective; infection rates rise with instar examined, overall infection prevalence rates are often 50% and may be much higher; as for mammals, mixed infections in bugs are common.[94,96,98] Although infection of the triatomine salivary glands is not described for *T. cruzi*, there is analogous intense infection of the rectum with abundant epimastigotes attached to the epithelium and free in the lumen, as well as metacyclic trypomastigotes. Salivary gland infections cannot be considered a prerequisite for

Downloaded From Vet-eBooks.com

genetic exchange to occur in triatomines—they are not a feature of the *Leishmania* life cycle in sand flies and many examples exist of genetic exchange in other protozoan flagellates in the gut lumen of insect species.[119] There are also prominent candidate physiological triggers in triatomines, for example, hormonal changes that govern molting, known to precipitate genetic exchange in flagellates of other insects, or starvation/nutritional depletion, which appears to encourage dispersive flight in male triatomine bugs.

Clearly, the occurrence of genetic exchange in *T. cruzi* during the vector stage of the life cycle deserves further investigation. With approximately 127 of the 140 or more triatomine species native to the Americas and at least six *T. cruzi* genetic lineages, there are multiple scenarios to be explored and indubitably new discoveries to be made, aided by the latest technological advances.

3.5 Behavior of Experimental Hybrids

The phenomenon of hybrid vigor (heterosis), in which hybrids tolerate unusually severe conditions or thrive and outcompete nonhybrids, is well known for a variety of organisms.[120] On the other hand, hybrids may have significantly reduced viability and be difficult to recover from experimental or natural populations. Hybrid vigor was explored theoretically for a naturally occurring hybrid *T. cruzi* by comparing the catalytic efficiency of the three individual glucose phosphate isomerase isoenzymes in hybrids, and against the individual isoenzymes in nonhybrids.[121] Although the isoenzymes clearly differed in temperature stabilities, their catalytic efficiencies appeared to be similar. Comprehensive phenotypic comparisons of natural TcV and TcVI hybrid strains with TcII and TcIII representative parental strains will be required to establish whether heterosis contributed to the success of TcV and TcVI in becoming established in domestic transmission cycles. In this context, it is clearly of interest to know whether experimentally derived hybrid *T. cruzi* clones have decreased or increased vigor. The experimental hybrid clones described by Gaunt et al.[9] grew vigorously in vitro, and in preliminary in vivo comparisons in immunocompromised SCID (severe combined immunodeficiency) mice, the hybrids readily established infections and produced abundant pseudocysts in heart and skeletal muscle; pseudocysts were also seen in smooth muscle of the alimentary tract.[122] The hybrid clones therefore appear to be at least as virulent as their parents. Genotype of the infecting strain or mixture of strains may have an important impact upon disease pathology in humans, for instance, as a result of differences in tissue tropism.[118,123] Results from experimental investigations of mixed infection dynamics also suggest that dual-clone mixtures of *T. cruzi* behave differently than would be expected simply based on the behavior of each clone individually, both in bugs,[124,125] and in animal models.[126–128] Imaging of transgenic bioluminescent biological clones of *T. cruzi* parents and derived hybrid progeny now provides a means of following their comparative behavior and virulence in vivo, throughout the acute and prolonged chronic phases of mouse model infections.[129]

Downloaded From Vet-eBooks.com

3.6 Mechanism of Genetic Exchange in Experimental and Natural Hybrids

The mechanism of genetic exchange that produced the six experimental hybrid *T. cruzi* clones is at least partially well understood. MLEE, karyotype analysis, random amplification of polymorphic DNA (RAPD), and multilocus microsatellite typing (MLMT) clearly showed that the hybrids had not inherited alleles in typical Mendelian ratios, that is, one allele per locus from each parent.[9] In fact, for most loci tested, the hybrids had inherited at least two alleles from both parents, although a minority of the parental alleles were absent. *Trypanosoma cruzi* hybrid clones each displayed one of the parental mitochondrial maxicircle genotypes but not both, as is reported (earlier) for *T. brucei*; the dynamics for inheritance of minicircles remains unexplored in *T. cruzi* but may well involve reassortment flow cytometric analysis of DNA content demonstrated that the hybrid clones were subtetraploid (about intermediate between 3N and 4N), and that DNA content was relatively stable after passage through mice.[122] However, following long-term in vitro culture, progressive and gradual decline in DNA content has been observed without any evidence of a meiotic reduction division, which would be expected to result in the halving of ploidy in a single step (Lewis et al., unpublished data). This situation is in contrast with the general consensus for the typical program of genetic exchange in *T. brucei*, which evidently does involve meiosis and Mendelian inheritance. Furthermore, analysis of natural *T. cruzi* hybrids (TcV and TcVI) showed that their genotypes were consistent with their being typical meiotic F1 progeny from a TcII × TcIII cross.[16] Nevertheless, more than one mechanism may occur; hybridization and genome erosion do have a biological precedent in the parasexual mechanism of genetic exchange in the pathogenic yeast *Candida albicans*.[130] In *C. albicans* the genome erosion following fusion of diploids is by random, concerted loss of whole chromosomes over the course of repeated mitotic replication and results in diploid or near-diploid recombinants.

The mechanism of genome erosion in *T. cruzi* requires more extensive comparative genotyping of hybrids and parents. The full extent of recombination and mosaic formation that took place, at chromosomal and intragenic levels, during the experimental *T. cruzi* cross is also not clear and requires further investigation. Both genetic exchange and genome erosion mechanisms should be revealed by comparative sequencing of the entire TcI parental and derived experimental progeny genomes, which is currently in progress (Talavares, Lewis et al., unpublished). In the context of these observations on *T. cruzi*, it is interesting to note that a proportion of polyploid progeny may also occur as products of genetic crossing experiments with *T. brucei* and *Leishmania.* As described earlier, haploid gametes of *T. brucei* were reported,[55] suggesting that in *T. brucei* triploid progeny may be the result of fusion prior to reduction division of gametes.

The interesting results of the DNA content analysis of the experimental cross, in particular, the elevated subtetraploid DNA content of the hybrids, led us to undertake a much wider study of DNA content among field isolates representing all six of the known *T. cruzi* genetic lineages. As described by Lewis et al.,[122] and in agreement with the pioneering observations by,[131] extraordinarily wide diversity in DNA content

Downloaded From Vet-eBooks.com

was observed (Fig. 20.3). Some trends were observed relating to the DNA content of cloned strains from the different genetic lineages, for example, TcI on average clearly had the lowest DNA content, with few outlying isolates, while TcII and TcIV presented a wide range of DNA contents. The key question, however, was whether naturally occurring hybrid TcV and TcVI strains had elevated DNA contents, compatible with the fusion of diploids, as deduced from the experimental cross. This proved not to be the case. TcV and TcVI had DNA contents equivalent to those of their parental TcII and TcIII lineages, implying diploidy. This was confirmed by genome-scale MLMT and DNA sequencing of nuclear and mitochondrial (maxicircle) genes in TcII, TcIII, TcV, and TcVI isolates, which indicated TcV and TcVI carried one TcII allele and one TcIII allele per locus, with only rare instances of allelic aneuploidy.[122] TcV and TcVI present very little intralineage diversity, but the extensive genotyping clearly shows they are distinct from each other, confirming the original MLEE-based distinction. TcV and TcVI thus resemble normal F1 meiotic progeny of hybridization events between TcII and TcIII that have undergone clonal expansion within the domestic niche—they are essentially diploid with fixed heterozygous genotypes comprising equal proportions of TcII and TcIII alleles. There are therefore

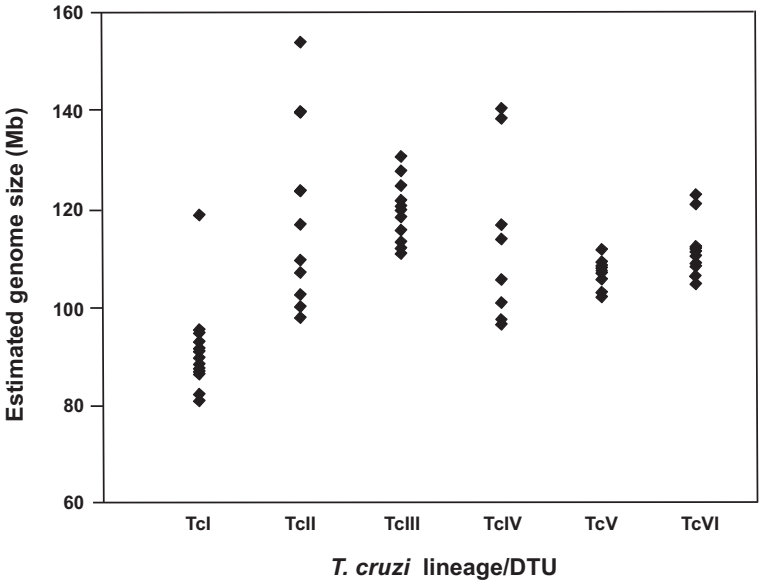


Figure 20.3 Variation in DNA content within and between *Trypanosoma cruzi* genetic lineages or DTUs. Each diamond indicates a different *T. cruzi* clone. Flow cytometry was used to measure the relative DNA content of parasites labeled with the fluorescent dye propidium iodide. Genome sizes were estimated based on the predicted size of the CL Brener genome.[161] Adapted from Lewis MD, Llewellyn MS, Gaunt MW, Yeo M, Carrasco HJ, Miles MA. Flow cytometric analysis and microsatellite genotyping reveal extensive DNA content variation in *Trypanosoma cruzi* populations and expose contrasts between natural and experimental hybrids. *Int J Parasitol* 2009;**39**:1305−17.

Downloaded From Vet-eBooks.com

contrasts with the aneuploid experimental hybrids. These could stem from mechanistic differences in interlineage, as opposed to intralineage, recombination. Alternatively, the physiological cues that might be required for meiotic reductive division may well have been absent under the conditions used for experimental crosses. It is not known how the diploid state was reached by TcV/VI and so operation of the fusion–erosion mechanism or other genetic exchange mechanisms in natural recombination events should not be ruled out at this stage.

The existence of TcV and TcVI clearly shows that natural recombination events have been a feature of the evolution of *T. cruzi*. Whether it is a contemporary phenomenon in active transmission cycles remains an open question. However there is evidence that recombination may be more frequent than previously thought. For example, molecular dating analyses indicated that the TcV and TcVI hybrids conservatively have an origin within the last 60,000 years.[16] Comparisons between the nuclear and mitochondrial genotypes of TcI, TcIII, and TcIV have revealed several striking instances of mitochondrial introgression.[13,92,132] Multilocus sequence typing (MLST) of reference strains has now shown that TcIII and TcIV could be descended from ancient hybridization between TcI and TcII,[15,92] although this has been questioned.[133] Furthermore, unexpectedly high levels of homozygosity within sylvatic populations of TcI and TcIII may be compatible with the occurrence of natural intralineage recombination.[132,134,135] Perhaps most importantly, genetics studies using more intensive sampling at the micro-scale of individual undisturbed transmission cycles have finally revealed evidence of sexually recombining *T. cruzi* populations.[136,137] In summary, as of 2016, recent research suggests that genetic exchange is more frequent among natural populations than so far described, with significant consequent epidemiological implications.[138–140] Messenger and Miles[132] provide a concise review of evidence for genetic exchange in natural populations of *T. cruzi*.

3.7 Implications for Epidemiology

The successful experimental *T. cruzi* genetic cross was a significant research milestone in that it proved this organism has an extant capacity for genetic exchange, with consequent epidemiological implications, and the unusual genetic mechanism continues to be of considerable fundamental interest. However, the main implications of genetic exchange in *T. cruzi* come not from examination of the experimental hybrids but from genetic analysis of natural populations. The hybrid DTUs, TcV and TcVI, have thrived and spread dramatically among human populations in Bolivia, Argentina, Chile, Paraguay, and the extreme south of Brazil, where severe clinical manifestations of Chagas disease are common, including chagasic cardiomyopathy, megaesophagus, megacolon, and congenital transmission[90,141] (Messenger, 2015 #3383). TcV and TcVI have probably been propagated in conjunction with the dispersion of *Triatoma infestans*, the principal domestic triatomine vector in the Southern Cone region, which was itself spread in association with the activities and migration of humans.[142,143] There is therefore a potentially great epidemiological risk attached to the emergence of new recombinant lineages, which would have unpredictable phenotypes, for example, with respect to pathogenicity or transmission potential. Furthermore, new

Downloaded From Vet-eBooks.com

high impact hybridization events cannot be excluded as the dynamics of sylvatic and domestic transmission cycles alter with development, environmental change, and migration of human populations.

4. *Leishmania*

4.1 Evidence of Genetic Exchange in Natural Populations

Numerous species of *Leishmania* have been named as agents of leishmaniasis. The multiple species of *Leishmania* and their complex taxonomy contrasts with *T. cruzi* as the sole designated species causing Chagas disease. This taxonomic contrast is despite the remarkable diversity of the *T. cruzi* genetic lineages, exceeding the level of diversity that is used to support definition of several distinct *Leishmania* species.[90] In fact, there has been a tendency to describe new *Leishmania* species based on minor genotypic differences; some species have been shown to be synonyms (e.g., *L. infantum* and *Leishmania chagasi*; *Leishmania donovani* and *L. archibaldi*) and several other species are of questionable validity.[144–146] Nevertheless, molecular methods have also revealed hitherto unexpected genetic diversity within species, for example, within *L. donovani*.

The two subgenera of *Leishmania*, subgenus *Leishmania* and subgenus *Viannia*, were initially distinguished based on differences in clinical presentation, and subsequently on the distribution of infection within the alimentary tract of the sand fly vectors. The *L. donovani* complex (subgenus *Leishmania; L. donovani* and *L. infantum*) causes human visceral leishmaniasis (VL, kala-azar), which is the most devastating form of human leishmaniasis. *Leishmania infantum* is also responsible for widespread severe canine visceral leishmaniasis (CVL). Symptomatic human VL is nearly always fatal in the absence of treatment, and an estimated 500,000 symptomatic cases occur annually. Several other species of the subgenus *Leishmania* are responsible for diverse clinical presentations of cutaneous disease in both the Old World (*L. major*, *Leishmania tropica*, *Leishmania aethiopica*) and the New World (*Leishmania amazonensis*, *Leishmania mexicana*), occasionally presenting with disseminated cutaneous lesions (*L. aethiopica*, *L. amazonensis*).

The subgenus *Viannia* is confined to the New World, where *Leishmania braziliensis* causes the most severe form of the cutaneous disease, mucocutaneous leishmaniasis (MCL), with potentially catastrophic metastatic invasion and destruction of the nasopharynx. *Leishmania peruviana* is very closely related to *L. braziliensis* but is more prevalent in higher altitude regions of Peru, where in contrast it is associated with simple, self-healing cutaneous lesions. Infections of *Leishmania guyanensis* and *Leishmania panamensis,* both within the subgenus *Viannia,* may spread beyond the initial site of infection but are usually less aggressive than *L. braziliensis*.

In parallel with research on *T. cruzi* diversity, a similar range of molecular methods has been applied to the genus *Leishmania*. Historically, as with *T. cruzi*, population structure of all *Leishmania* has been considered to be fundamentally clonal. However, evidence for genetic exchange in natural populations of

Downloaded From Vet-eBooks.com

Leishmania has repeatedly emerged in the form of both interspecific and intraspecific hybrids bearing recombinant genotypes. For example, natural hybrids have been described between *L. braziliensis* and *L. panamensis* in Nicaragua,[17] between *L. braziliensis* and *L. guyanensis* in Venezuela,[147] and between *L. braziliensis* and *L. peruviana* in Peru.[148] In the latter case, the *L. braziliensis/L. peruviana* hybrids were highly prevalent among patients, and occurred in some patients with mucosal disease. Furthermore, nine hybrid genotypes were discovered in a single Peruvian endemic region.[149] In the Old World, *L. major/Leishmania arabica* hybrids have been described,[18,150] and *L. major/L. donovani* hybrids (originally designated *L. major/L. infantum* hybrids) from patients with HIV infection.[19] Putative parental and hybrid phenotypes of the *L. donovani* complex (*L. donovani;* "*L. archibaldi*") occur sympatrically in East Africa, and sequencing of housekeeping genes encoding enzymes shows mosaic characters across such strains.[151] A widespread lineage of *L. tropica* appears to be disseminated from a recent recombination event.[152] Inbreeding was detected in natural populations of both *L. braziliensis* and *L. guyanensis* based on linkage disequilibrium (LD), with a deficit of heterozygosity.[153] Comparative genome sequencing to detect recombination showed evidence of hybridization and subsequent selfing among vector-isolated *L. infantum* from Southeastern Turkey.[154] Multilocus sequence typing and microsatellites confirmed the presence of sympatric putative parents and hybrid progeny in an endemic focus of VL in Ethiopia.[155] In heroic comparative genomics of 204 strains of *L. donovani* from the Indian subcontinent, Imamura et al. (2016) found evidence of genetic exchange, associated with the spread of the LdAQP1 mutation marker of resistance to antimonials. Thus, there is an accumulation of evidence of intra- and interspecific genetic exchange in natural populations of *Leishmania*.

4.2 Genetic Crosses

Akopyants et al.[6] finally demonstrated recombination in *Leishmania* experimentally. They cotransmitted pairs of transgenic *L. major* strains resistant to different selective drugs through the natural sand fly vector, *Phlebotomus duboscqi*, and recovered parasites that were resistant to both drugs. Nine double drug−resistant populations were recovered directly from coinfected sand flies and two from mice bitten by such flies; these yielded a total of 18 progeny clones. Genetic analysis of homozygous parental markers, assigned to several chromosomes, showed that all clones carried both parental alleles, proving they were hybrid. Maxicircle kDNA genotypes were uniparentally inherited. Phenotypic analysis also demonstrated clear segregation of dominant traits in different hybrid progeny. The genotypes were consistent with canonical meiosis generating heterozygous F1 progeny. However, 7 of the 18 clones were triploid. Experimental crosses using parental clones carrying red and green fluorescent markers in conjunction with separation of single cells by fluorescence-activated cell sorting revealed the presence of "yellow" hybrid cells with the expression of both fluorescent markers.[156]

Subsequently, it was shown that genetic crosses, in both the natural vector *P. duboscqi* and a nonnatural permissive vector *Leishmania longipalpis*, could be

Downloaded From Vet-eBooks.com

performed with facility and between geographically diverse *L. major* strains, with similar efficiencies in all pairwise crosses.[157] Of 96 derived clonal lines, the majority were diploid but some were triploid or tetraploid. A low frequency of uniparental allelic inheritance, more so in the nonnatural vector species, was attributed to loss of heterozygosity. Again, only uniparental maxicircle inheritance occurred. No distinct male or female gametes have been described, but of the developmental stages in the sand fly (procyclic, nectomonad, haptomonad, and metacyclic), the nectomonad stage is considered to be the most likely stage at which genetic exchange takes place.[157]

Consistent with the observation of *L. major*/*L. donovani* hybrids in natural populations,[19] Romano et al.[4] achieved interspecies crosses of *L. major* and *L. infantum* in the permissive vector *L. longipalpis*, yielding hybrid progeny that displayed similar patterns of inheritance and ploidy among hybrids from the *L. major* crosses. Most importantly, trait segregation of cutaneous or visceral dissemination and pathology was seen among the progeny. Finally, Calvo-Alvarez et al.[158] showed that intraclonal crosses of *L. infantum* in a natural vector *Phlebotomus perniciosus* could produce a viable hybrid, somewhat less virulent in a mouse model than the parental clone. Many hybrid *L. donovani* clones have recently been generated experimentally in other sand fly species (Sadlova et al., unpublished data as of mid-2016).

4.3 Implications for Epidemiology

Research progress on genetic exchange in *Leishmania* and in trypanosomes has now been dramatic and far reaching, albeit much remains to be done to understand precise alternative mechanisms. *L. major*/*L. donovani* hybrids have increased transmission potential, since they were transmitted efficiently by the otherwise *L. major*–specific vector, *Phlebotomus papatasi*.[159,160] Such results highlight the potential epidemiological impact of genetic exchange in *Leishmania*: hybrids may have the disturbing potential to expand transmission of visceral disease and invade new geographical regions.

Thus, the fact that *Leishmania* spp. can undergo genetic exchange is of profound epidemiological importance.[90] Genetic exchange has implications for heterosis (hybrid vigor), the emergence and spread of virulent strains, resistance to chemotherapeutics, exploitation of different vectors and hosts, and adaptation to new ecological niches that may provide a selective advantage. As with *T. brucei* and *T. cruzi*, virulent clones may emerge and predominate in some epidemiologically important populations. However, the perception that genetic exchange is rare and of little evolutionary or extant epidemiological consequence is no longer tenable.

Abbreviations

MLEE	Multilocus enzyme electrophoresis
MLMT	Multilocus microsatellite typing
MLST	Multilocus sequence typing

Downloaded From Vet-eBooks.com

Acknowledgments

We thank the Wellcome Trust, the European Union Seventh Framework Programme, contract number 223034 (ChagasEpiNet) for financial support. We especially thank many current and previous research collaborators working on African trypanosomiasis, *Trypanosoma cruz*i and Chagas disease, and leishmaniasis, particularly those in Africa and Latin America.

References

1. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, et al. The genome of the African trypanosome *Trypanosoma brucei*. *Science* 2005;**309**:416−22.
2. El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, et al. Comparative genomics of trypanosomatid parasitic protozoa. *Science* 2005;**309**:404−9.
3. Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, et al. The genome of the kinetoplastid parasite, *Leishmania major*. *Science* 2005;**309**:436−42.
4. Romano A, Inbar E, Debrabant A, Charmoy M, Lawyer P, Ribeiro-Gomes F, et al. Cross-species genetic exchange between visceral and cutaneous strains of *Leishmania* in the sand fly vector. *Proc Natl Acad Sci USA* 2014;**111**:16808−13.
5. Schmid-Hempel R, Salathe R, Tognazzo M, Schmid-Hempel P. Genetic exchange and emergence of novel strains in directly transmitted trypanosomatids. *Infect Genet Evol* 2011;**11**:564−71.
6. Akopyants NS, Kimblin N, Secundino N, Patrick R, Peters N, Lawyer P, et al. Demonstration of genetic exchange during cyclical development of *Leishmania* in the sand fly vector. *Science* 2009;**324**:265−8.
7. Jenni L, Marti S, Schweizer J, Betschart B, Lepage RWF, Wells JM, et al. Hybrid formation between African trypanosomes during cyclical transmission. *Nature* 1986;**322**:173−5.
8. MacLeod A, Tweedie A, McLellan S, Taylor S, Cooper A, Sweeney L, et al. Allelic segregation and independent assortment in *Trypanosoma brucei* crosses: proof that the genetic system is Mendelian and involves meiosis. *Mol Biochem Parasitol* 2005;**143**:12−9.
9. Gaunt MW, Yeo M, Frame IA, Stothard JR, Carrasco HJ, Taylor MC, et al. Mechanism of genetic exchange in American trypanosomes. *Nature* 2003;**421**:936−9.
10. Tait A. Evidence for diploidy and mating in trypanosomes. *Nature* 1980;**287**:536−8.
11. Tibayrenc M, Kjellberg F, Ayala FJ. A clonal theory of parasitic protozoa: the population structures of *Entamoeba, Giardia, Leishmania, Naegleria, Plasmodium, Trichomonas* and *Trypanosoma* and their medical and taxonomical consequences. *Proc Natl Acad Sci USA* 1990;**87**:2414−8.
12. Carrasco HJ, Frame IA, Valente SA, Miles MA. Genetic exchange as a possible source of genomic diversity in sylvatic populations of *Trypanosoma cruzi*. *Am J Trop Med Hyg* 1996;**54**:418−24.
13. Machado CA, Ayala FJ. Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*. *Proc Natl Acad Sci USA* 2001;**98**:7396−401.
14. Sturm NR, Vargas NS, Westenberger SJ, Zingales B, Campbell DA. Evidence for multiple hybrid groups in *Trypanosoma cruzi*. *Int J Parasitol* 2003;**33**:269−79.

Downloaded From Vet-eBooks.com

15. Westenberger SJ, Barnabe C, Campbell DA, Sturm NR. Two hybridization events define the population structure of *Trypanosoma cruzi*. *Genetics* 2005;**171**:527−43.
16. Lewis MD, Llewellyn MS, Yeo M, Acosta N, Gaunt MW, Miles MA. Recent, independent and anthropogenic origins of *Trypanosoma cruzi* hybrids. *PLoS NTD* 2011;**5**:e1363.
17. Belli AA, Miles MA, Kelly JM. A putative *Leishmania panamensis* − *Leishmania braziliensis* hybrid is a causative agent of human cutaneous leishmaniaisis in Nicaragua. *Parasitology* 1994;**109**:435−42.
18. Kelly JM, Law JM, Chapman CJ, Van Eyes GJJM, Evans DA. Evidence of genetic recombination in *Leishmania*. *Mol Biochem Parasitol* 1991;**46**:253−64.
19. Ravel C, Cortes S, Pratlong F, Morio F, Dedet JP, Campino L. First report of genetic hybrids between two very divergent *Leishmania* species: *Leishmania infantum* and *Leishmania major*. *Int J Parasitol* 2006;**36**:1383−8.
20. Gibson WC, de C Marshall TF, Godfrey DG. Numerical analysis of enzyme polymorphism: a new approach to the epidemiology and taxonomy of trypanosomes of the subgenus *Trypanozoon*. *Adv Parasitol* 1980;**18**:175−246.
21. Maudlin I. Transmission of African Trypanosomiasis: interactions among tsetse immune system, symbionts and parasites. *Adv Dis Vector Res* 1991;**7**:117−48.
22. Eid J, Sollner-Webb B. Stable integrative transformation of *Trypanosoma brucei* that occurs exclusively by homologous recombination. *Proc Natl Acad Sci USA* 1991;**88**: 2118−21.
23. Lee G-SM, Van der Ploeg LHT. Homologous recombination and stable transfection in the parasitic protozoan *Trypanosoma brucei*. *Science* 1990;**250**:1583−7.
24. Ten Asbroek ALMA, Ouellette M, Borst P. Targeted insertion of the neomycin phosphotransferase gene into the tubulin gene cluster of *Trypanosoma brucei*. *Nature* 1990; **348**:174−5.
25. Gibson W, Whittington H. Genetic exchange in *Trypanosoma brucei*: selection of hybrid trypanosomes by introduction of genes conferring drug resistance. *Mol Biochem Parasitol* 1993;**60**:19−26.
26. Gibson W, Peacock L, Ferris V, Williams K, Bailey M. The use of yellow fluorescent hybrids to indicate mating in *Trypanosoma brucei*. *Parasit Vector* 2008;**1**:4.
27. Gibson WC. Will the real *Trypanosoma brucei gambiense* please stand up? *Parasitol Today* 1986;**2**:255−7.
28. Dukes P, Kaukus A, Hudson KM, Asonganyi T, Gashumba JK. A new method for isolating *Trypanosoma brucei gambiense* from sleeping sickness patients. *Trans Roy Soc Trop Med Hyg* 1989;**83**:636−9.
29. Bingle LEH, Eastlake JL, Bailey M, Gibson WC. A novel GFP approach for the analysis of genetic exchange in trypanosomes allowing the in situ detection of mating events. *Microbiology* 2001;**147**:3231−40.
30. Gibson W, Winters K, Mizen G, Kearns J, Bailey M. Intraclonal mating in *Trypanosoma brucei* is associated with out-crossing. *Microbiology* 1997;**143**:909−20.
31. Degen R, Pospichal H, Enyaru J, Jenni L. Sexual compatibility among *Trypanosoma brucei* isolates from an epidemic area in southeastern Uganda. *Parasitol Res* 1995;**81**: 253−7.
32. Gibson WC, Garside LH. Genetic exchange in *Trypanosoma brucei brucei*: variable location of housekeeping genes in different trypanosome stocks. *Mol Biochem Parasitol* 1991;**45**:77−90.
33. Schweizer J, Pospichal H, Hide G, Buchanan N, Tait A, Jenni L. Analysis of a new genetic cross between 2 East African *Trypanosoma brucei* clones. *Parasitology* 1994;**109**:83−93.

Downloaded From Vet-eBooks.com

34. Turner CMR, Sternberg J, Buchanan N, Smith E, Hide G, Tait A. Evidence that the mechanism of gene exchange in *Trypanosoma brucei* involves meiosis and syngamy. *Parasitology* 1990;**101**:377−86.
35. Gibson W, Peacock L, Ferris V, Fischer K, Livingstone J, Thomas J, et al. Genetic recombination between human and animal parasites creates novel strains of human pathogen. *PLoS NTD* 2015;**9**:e0003665.
36. Gibson WC. Analysis of a genetic cross between *Trypanosoma brucei rhodesiense* and *T. b. brucei*. *Parasitology* 1989;**99**:391−402.
37. Gibson WC, Mizen VH. Heritability of the trait for human infectivity in genetic crosses of *Trypanosoma brucei* ssp. *Trans Roy Soc Trop Med Hyg* 1997;**91**:236−7.
38. Turner CMR, McLellan S, Lindergard LAG, Bisoni L, Tait A, MacLeod A. Human infectivity trait in *Trypanosoma brucei*: stability, heritability and relationship to *sra* expression. *Parasitology* 2004;**129**:445−54.
39. De Greef C, Imberechts H, Matthyssons G, Van Meirvenne N, Hamers R. A gene expressed only in serum-resistant variants of *Trypanosoma brucei rhodesiense*. *Mol Biochem Parasitol* 1989;**36**:169−76.
40. Xong VH, Vanhamme L, Chamekh M, Chimfwembe CE, Van den Abbeele J, Pays A, et al. A VSG expression site-associated gene confers resistance to human serum in *Trypanosoma rhodesiense*. *Cell* 1998;**95**:839−46.
41. Evans DA, Ellis DE. Recent observations on the behaviour of certain trypanosomes within their insect hosts. *Adv Parasitol* 1983;**22**:1−42.
42. Schweizer J, Jenni L. Hybrid formation in the lifecycle of *Trypanosoma* (*T.*) *brucei*: detection of hybrid trypanosomes in a midgut-derived isolate. *Acta Trop* 1991;**48**: 319−21.
43. Schweizer J, Pospichal H, Jenni L. Hybrid formation between African trypanosomes in vitro. *Acta Trop* 1991;**49**:237−40.
44. Schweizer J, Tait A, Jenni L. The timing and frequency of hybrid formation in African trypanosomes during cyclical transmission. *Parasitol Res* 1988;**75**:98−101.
45. Gibson W, Bailey M. Genetic exchange in *Trypanosoma brucei*: evidence for meiosis from analysis of a cross between drug resistant transformants. *Mol Biochem Parasitol* 1994;**64**:241−52.
46. Gibson W, Stevens J. Genetic exchange in the Trypanosomatidae. *Adv Parasitol* 1999;**43**: 1−46.
47. Peacock L, Ferris V, Bailey M, Gibson W. Dynamics of infection and competition between two strains of *Trypanosoma brucei brucei* in the tsetse fly observed using fluorescent markers. *Kinetoplastid Biol Dis* 2007;**6**:4.
48. Gottesdiener K, Garcia-Anoveros J, Lee G-SM, Van der Ploeg LHT. Chromosome organization of the protozoan *Trypanosoma brucei*. *Mol Cell Biol* 1990;**10**:6079−83.
49. Melville SE, Leech V, Gerrard CS, Tait A, Blackwell JM. The molecular karyotype of the megabase chromosomes of *Trypanosoma brucei* and the assignment of chromosome markers. *Mol Biochem Parasitol* 1998;**94**:155−73.
50. Sternberg J, Turner CMR, Wells JM, Ranford-Cartwright LC, Lepage RWF, Tait A. Gene exchange in African trypanosomes: frequency and allelic segregation. *Mol Biochem Parasitol* 1989;**34**:269−80.
51. Cooper A, Tait A, Sweeney L, Tweedie A, Morrison L, Turner CMR, et al. Genetic analysis of the human infective trypanosome, *Trypanosoma brucei gambiense*: chromosomal segregation, crossing over and the construction of a genetic map. *Genome Biol* 2008;**9**.

Downloaded From Vet-eBooks.com

52. MacLeod A, Tweedie A, McLellan S, Taylor S, Hall N, Berriman M, et al. The genetic map and comparative analysis with the physical map of *Trypanosoma brucei*. *Nucl Acids Res* 2005;**33**:6688−93.
53. Ramesh MA, Malik SB, Logsdon JM. A phylogenomic inventory of meiotic genes: evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. *Curr Biol* 2005;**15**: 185−91.
54. Peacock L, Ferris V, Sharma R, Sunter J, Bailey M, Carrington M, et al. Identification of the meiotic life cycle stage of *Trypanosoma brucei* in the tsetse fly. *Proc Natl Acad Sci USA* 2011;**108**:3671−6.
55. Peacock L, Bailey M, Carrington M, Gibson W. Meiosis and haploid gametes in the pathogen *Trypanosoma brucei*. *Curr Biol* 2014;**24**:1−6.
56. Kooy RF, Hirumi H, Moloo SK, Nantuyla VM, Dukes P, Van der Linden PM, et al. Evidence for diploidy in metacyclic forms of African trypanosomes. *Proc Natl Acad Sci USA* 1989;**86**:5469−72.
57. Shapiro SZ, Naessens J, Liesegang B, Moloo SK, Magondu J. Analysis by flow cytometry of DNA synthesis during the life cycle of African trypanosomes. *Acta Trop* 1984;**41**: 313−23.
58. Zampetti-Bosseler F, Schweizer J, Pays E, Jenni L, Steinert M. Evidence for haploidy in metacyclic forms of *Trypanosoma brucei*. *Proc Natl Acad Sci USA* 1986;**83**:6063−4.
59. Gibson W, Garside L, Bailey M. Trisomy and chromosome size changes in hybrid trypanosomes from a genetic cross between *Trypanosoma brucei rhodesiense* and *T. b. brucei*. *Mol Biochem Parasitol* 1992;**52**:189−200.
60. Gibson W, Kanmogne G, Bailey M. A successful backcross in *Trypanosoma brucei*. *Mol Biochem Parasitol* 1995;**69**:101−10.
61. Paindavoine P, Zampetti-Bosseler F, Pays E, Schweizer J, Guyaux M, Jenni L, et al. Trypanosome hybrids generated in tsetse flies by nuclear fusion. *EMBO J* 1986;**5**:3631−6.
62. Wells JM, Prospero TD, Jenni L, Le Page RWF. DNA contents and molecular karyotypes of hybrid *Trypanosoma brucei*. *Mol Biochem Parasitol* 1987;**24**:103−16.
63. Tait A, Buchanan N, Hide G, Turner M. Self-fertilisation in *Trypanosoma brucei*. *Mol Biochem Parasitol* 1996;**76**:31−42.
64. Sternberg J, Tait A, Haley S, Wells JM, Lepage RWF, Schweizer J, et al. Gene exchange in African trypanosomes: characterisation of a new hybrid genotype. *Mol Biochem Parasitol* 1988;**27**:191−200.
65. Peacock L, Ferris V, Bailey M, Gibson W. Intraclonal mating occurs during tsetse transmission of *Trypanosoma brucei*. *Parasit Vector* 2009;**2**:43.
66. Peacock L, Ferris V, Bailey M, Gibson W. Mating compatibility in the parasitic protist *Trypanosoma brucei*. *Parasit Vector* 2014;**7**:78.
67. Shapiro TA, Englund PT. The structure and replication of kinetoplast DNA. *Annu Rev Microbiol* 1995;**49**:117−43.
68. Gibson W, Crow M, Kearns J. Kinetoplast DNA minicircles are inherited from both parents in genetic crosses of *Trypanosoma brucei*. *Parasitol Res* 1997;**83**:483−8.
69. Gibson W, Garside L. Kinetoplast DNA mini-circles are inherited from both parents in genetic hybrids of *Trypanosoma brucei*. *Mol Biochem Parasitol* 1990;**42**:45−54.
70. Turner CMR, Hide G, Buchanan N, Tait A. *Trypanosoma brucei* − inheritance of kinetoplast DNA maxicircles in a genetic cross and their segregation during vegetative growth. *Expt Parasitol* 1995;**80**:234−41.
71. Cibulskis RE. Genetic variation in *Trypanosoma brucei* and the epidemiology of sleeping sickness in the Lambwe Valley, Kenya. *Parasitology* 1992;**104**:99−109.

Downloaded From Vet-eBooks.com

72. Stevens JR, Welburn SC. Genetic processes within an epidemic of sleeping sickness in Uganda. *Parasitol Res* 1993;**79**:421−7.
73. Hide G, Welburn SC, Tait A, Maudlin I. Epidemiological relationships of *Trypanosoma brucei* stocks from South East Uganda: evidence for different population structures in human infective and non-infective isolates. *Parasitology* 1994;**109**:95−111.
74. Stevens JR, Tibayrenc M. *Trypanosoma brucei* s.l.: evolution, linkage and the clonality debate. *Parasitology* 1996;**112**:481−8.
75. Wijers DJB. Factors that may influence the infection rate of *Glossina palpalis* with *Trypanosoma gambiense*. I. The age of the fly at the time of the infected feed. *Ann Trop Med Parasit* 1958;**52**:385−90.
76. Truc P, Jamonneau V, NGuessan P, NDri L, Diallo PB, Cuny G. *Trypanosoma brucei* ssp. and *T. congolense*: mixed human infection in Cote d'Ivoire. *Trans Roy Soc Trop Med Hyg* 1998;**92**:537−8.
77. Balmer O, Tanner M. Prevalence and implications of multiple-strain infections. *Lancet Infect Dis* 2011;**11**:868−78.
78. Godfrey DG, Killick-Kendrick R. Bovine trypanosomiasis in Nigeria. I. The inoculation of blood into rats as a method of survey in the Donga Valley, Benue province. *Ann Trop Med Parasit* 1961;**55**:287−92.
79. Nyeko JHP, Ole-Moiyoi OK, Majiwa P, Otieno LH, Ociba PM. Characterisation of trypanosome isolates from cattle in Uganda using species-specific DNA probes reveals predominance of mixed infections. *Insect Sci Its Appl* 1990;**11**:271−80.
80. Adams ER, Malele II, Msangi AR, Gibson WC. Trypanosome identification in wild tsetse populations in Tanzania using generic primers to amplify the ribosomal RNA ITS-1 region. *Acta Trop* 2006;**100**:103−9.
81. Lehane MJ, Msangi AR, Whitaker CJ, Lehane SM. Grouping of trypanosome species in mixed infections in *Glossina pallidipes*. *Parasitology* 2000;**120**:583−92.
82. MacLeod A, Turner CMR, Tait A. A high level of mixed *Trypanosoma brucei* infections in tsetse flies detected by three hypervariable minisatellites. *Mol Biochem Parasitol* 1999; **102**:237−48.
83. Masiga DK, McNamara JJ, Laveissiere C, Truc P, Gibson WC. A high prevalence of mixed trypanosome infections in tsetse flies in Sinfra, Cote d'Ivoire, detected by DNA amplification. *Parasitology* 1996;**112**:75−80.
84. Enyaru JC, Stevens JR, Odiit M, Okuna NM, Carasco JF. Isoenzyme comparison of *Trypanozoon* isolates from two sleeping sickness areas of south-eastern Uganda. *Acta Trop* 1993;**55**:97−115.
85. Gibson WC, Gashumba JK. Isoenzyme characterisation of some *Trypanozoon* stocks from a recent trypanosomiasis epidemic in Uganda. *Trans Roy Soc Trop Med Hyg* 1983;**77**: 114−8.
86. Gibson WC, Wellde BT. Characterisation of *Trypanozoon* stocks from the south Nyanza sleeping sickness focus in Western Kenya. *Trans Roy Soc Trop Med Hyg* 1985;**79**: 671−6.
87. Komba EK, Kibona SN, Ambwene AK, Stevens JR, Gibson WC. Genetic diversity among *Trypanosoma brucei rhodesiense* isolates from Tanzania. *Parasitology* 1997;**115**:571−9.
88. Echodu R, Sistrom M, Bateta R, Murilla G, Okedi L, Aksoy S, et al. Genetic diversity and population structure of *Trypanosoma brucei* in Uganda: implications for the epidemiology of sleeping sickness and nagana. *PLoS NTD* 2015;**9**:e0003353.
89. Balmer O, Beadell JS, Gibson W, Caccone A. Phylogeography and taxonomy of *Trypanosoma brucei*. *PLoS NTD* 2011;**5**:e961.

Downloaded From Vet-eBooks.com

90. Miles MA, Llewellyn MS, Lewis MD, Yeo M, Baleela R, Fitzpatrick S, et al. The molecular epidemiology and phylogeography of *Trypanosoma cruzi* and parallel research on *Leishmania*: looking back and to the future. *Parasitology* 2009;**136**:1509−28.
91. Messenger LA, Yeo M, Lewis MD, Llewellyn MS, Miles MA. Molecular genotyping of *Trypanosoma cruzi* for lineage assignment and population genetics. *Methods Mol Biol* 2015;**1201**:297−337.
92. Messenger LA, Miles MA, Bern C. Between a bug and a hard place: *Trypanosoma cruzi* genetic diversity and the clinical outcomes of Chagas disease. *Exp Rev Anti-Inf Ther* 2015; **13**:995−1029.
93. Zingales B, Andrade SG, Briones MRS, Campbell DA, Chiari E, Fernandes O, et al. A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. *Mem Oswaldo Cruz* 2009;**104**:1051−4.
94. Bosseno MF, Telleria J, Vargas F, Yaksic N, Noireau F, Morin A, et al. *Trypanosoma cruzi*: study of the distribution of two widespread clonal genotypes in Bolivian *Triatoma infestans* vectors shows a high frequency of mixed infections. *Expt Parasitol* 1996;**83**: 275−82.
95. Burgos JM, Begher S, Silva HMV, Bisio M, Duffy T, Levin MJ, et al. Case report: molecular identification of *Trypanosoma cruzi* I tropism for central nervous system in Chagas reactivation due to AIDS. *Am J Trop Med Hyg* 2008;**78**:294−7.
96. Cardinal MV, Lauricella MA, Ceballos LA, Lanati L, Marcet PL, Levin MJ, et al. Molecular epidemiology of domestic and sylvatic *Trypanosoma cruzi* infection in rural northwestern Argentina. *Int J Parasitol* 2008;**38**:1533−43.
97. Yeo M, Acosta N, Llewellyn M, Sanchez H, Adamson S, Miles GAJ, et al. Origins of Chagas disease: *Didelphis* species are natural hosts of *Trypanosoma cruzi* I and armadillos hosts of *Trypanosoma cruzi* II, including hybrids. *Int J Parasitol* 2005;**35**:225−33.
98. Yeo M, Lewis MD, Carrasco HJ, Acosta N, Llewellyn M, Valente SAD, et al. Resolution of multiclonal infections of *Trypanosoma cruzi* from naturally infected triatomine bugs and from experimentally infected mice by direct plating on a sensitive solid medium. *Int J Parasitol* 2007;**37**:111−20.
99. Barnabe C, Brisse S, Tibayrenc M. Population structure and genetic typing of *Trypanosoma cruzi*, the agent of Chagas disease: a multilocus enzyme electrophoresis approach. *Parasitology* 2000;**120**:513−26.
100. Fernandes O, Sturm NR, Derre R, Campbell DA. The mini-exon gene: a genetic marker for zymodeme III of *Trypanosoma cruzi*. *Mol Biochem Parasitol* 1998;**95**:129−33.
101. Zingales B, Souto RP, Mangia RH, Lisboa CV, Campbell DA, Coura JR, et al. Molecular epidemiology of American trypanosomiasis in Brazil based on dimorphisms of rRNA and mini-exon gene sequences. *Int J Parasitol* 1998;**28**:105−12.
102. Miles MA, Povoa MM, Prata A, Cedillos RA, Desouza AA, Macedo V. Do radically dissimilar *Trypanosoma cruzi* strains (zymodemes) cause Venezuelan and Brazilian forms of Chagas disease? *Lancet* 1981;**1**:1338−40.
103. Miles MA, Souza A, Povoa M, Shaw JJ, Lainson R, Toye PJ. Isozymic heterogeneity of *Trypanosoma cruzi* in first autochthonous patients with Chagas disease in Amazonian Brazil. *Nature* 1978;**272**:819−21.
104. Lisboa CV, Mangia RH, De Lima NRC, Martins A, Dietz J, Baker AJ, et al. Distinct patterns of *Trypanosoma cruzi* infection in *Leontopithecus rosalia* in distinct Atlantic coastal rainforest fragments in Rio de Janeiro-Brazil. *Parasitology* 2004;**129**:703−11.
105. Bhattacharyya T, Mills EA, Jansen AM, Miles MA. Prospects for *T. cruzi* lineage-specific serological surveillance of wild mammals. *Acta Trop* 2015;**151**:182−6.

Downloaded From Vet-eBooks.com

106. Carrasco HJ, Segovia M, Llewellyn MS, Morocoima A, Urdaneta-Morales S, Martinez C, et al. Geographical distribution of *Trypanosoma cruzi* genotypes in Venezuela. *PLoS NTD* 2012;**6**.
107. Tibayrenc M, Ayala FJ. The clonal theory of parasitic protozoa: 12 years on. *Trends Parasitol* 2002;**18**:405−10.
108. Miles MA. Ploidy, 'heterozygosity' and antigenic expression of South American trypanosomes. *Parasitologia* 1985;**27**:87−104.
109. Brisse S, Henriksson J, Barnabe C, Douzery EJP, Berkvens D, Serrano M, et al. Evidence for genetic exchange and hybridization in *Trypanosoma cruzi* based on nucleotide sequences and molecular karyotype. *Infect Genet Evol* 2003;**2**:173−83.
110. Weatherly DB, Boehlke C, Tarleton RL. Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. *BMC Genomics* 2009;**10**.
111. Franzen O, Talavera-Lopez C, Ochaya S, Butler CE, Messenger LA, Lewis MD, et al. Comparative genomic analysis of human infective *Trypanosoma cruzi* lineages with the bat-restricted subspecies *T. cruzi marinkellei*. *BMC Genomics* 2012;**13**.
112. Franzen O, Ochaya S, Sherwood E, Lewis MD, Llewellyn MS, Miles MA, et al. Shotgun sequencing analysis of *Trypanosoma cruzi* I Sylvio X10/1 and comparison with *T. cruzi* VI CL Brener. *PLoS NTD* 2011;**5**.
113. Stothard JR, Frame IA, Miles MA. Genetic diversity and genetic exchange in *Trypanosoma cruzi*: dual drug-resistant "progeny" from episomal transformants. *Mem Oswaldo Cruz* 1999;**94**:189−93.
114. Guevara P, Dias M, Rojas A, Crisante G, Abreu-Blanco MT, Umezara E, et al. Expression of fluorescent genes in *Trypanosoma cruzi* and *Trypanosoma rangeli* (Kinetoplastida: Trypanosomatidae) its application to parasite-vector biology. *J Med Entomol* 2005;**42**: 48−56.
115. Pires SF, DaRocha WD, Freitas JM, Oliveira LA, Kitten GT, Machado CR, et al. Cell culture and animal infection with distinct *Trypanosoma cruzi* strains expressing red and green fluorescent proteins. *Int J Parasitol* 2008;**38**:289−97.
116. Breniere SF, Morochi W, Bosseno MF, Ordonez J, Gutierrez T, Vargas F, et al. *Trypanosoma cruzi* genotypes associated with domestic *Triatoma sordida* in Bolivia. *Acta Trop* 1998;**71**:269−83.
117. Fernandes O, Mangia RH, Lisboa CV, Pinho AP, Morel CM, Zingales B, et al. The complexity of the sylvatic cycle of *Trypanosoma cruzi* in Rio de Janeiro state (Brazil) revealed by the non-transcribed spacer of the mini-exon gene. *Parasitology* 1999;**118**: 161−6.
118. Vago AR, Andrade LO, Leite AA, Reis DD, Macedo AM, Adad SJ, et al. Genetic characterization of *Trypanosoma cruzi* directly from tissues of patients with chronic Chagas disease − differential distribution of genetic types into diverse organs. *Am J Pathol* 2000; **156**:1805−9.
119. Sleigh MA. *The biology of Protozoa*. London: Edward Arnold; 1973.
120. Mallet J. Hybrid speciation. *Nature* 2007;**446**:279−83.
121. Widmer G, Dvorak JA, Miles MA. Temperature modulation of growth rates and glucosephosphate isomerase isozyme activity in Trypanosoma cruzi. *Mol Biochem Parasitol* 1987;**23**:55−62.
122. Lewis MD, Llewellyn MS, Gaunt MW, Yeo M, Carrasco HJ, Miles MA. Flow cytometric analysis and microsatellite genotyping reveal extensive DNA content variation in *Trypanosoma cruzi* populations and expose contrasts between natural and experimental hybrids. *Int J Parasitol* 2009;**39**:1305−17.

Downloaded From Vet-eBooks.com

123. Vago AR, Macedo AM, Oliveira RP, Andrade LO, Chiari E, Galvao LMC, et al. Kinetoplast DNA signatures of *Trypanosoma cruzi* strains obtained directly from infected tissues. *Am J Pathol* 1996;**149**:2153−9.
124. Pinto AD, de Lana M, Britto C, Bastrenta B, Tibayrenc M. Experimental *Trypanosoma cruzi* biclonal infection in *Triatoma infestans*: detection of distinct clonal genotypes using kinetoplast DNA probes. *Int J Parasitol* 2000;**30**:843−8.
125. Pinto AD, de Lana M, Bastrenta B, Barnabe C, Quesney V, Noel S, et al. Compared vectorial transmissibility of pure and mixed clonal genotypes of *Trypanosoma cruzi* in *Triatoma infestans*. *Parasitol Res* 1998;**84**:348−53.
126. de Lana M, Pinto AD, Bastrenta B, Barnabe C, Noel S, Tibayrenc M. *Trypanosoma cruzi*: infectivity of clonal genotype infections in acute and chronic phases in mice. *Expt Parasitol* 2000;**96**:61−6.
127. Martins HR, Silva RM, Valadares HMS, Toledo MJO, Veloso VM, Vitelli-Avelar DM, et al. Impact of dual infections on chemotherapeutic efficacy in BALB/c mice infected with major genotypes of *Trypanosoma cruzi*. *Antimicrob Agents Chemo* 2007;**51**:3282−9.
128. Martins HR, Toledo MJO, Veloso VM, Carneiro CM, Machado-Coelho GLL, Tafuri WL, et al. *Trypanosoma cruzi*: impact of dual-clone infections on parasite biological properties in BALB/c mice. *Expt Parasitol* 2006;**112**:237−46.
129. Lewis MD, Fortes Francisco A, Taylor MC, Burrell-Saward H, McLatchie AP, Miles MA, et al. Bioluminescence imaging of chronic *Trypanosoma cruzi* infections reveals tissue-specific parasite dynamics and heart disease in the absence of locally persistent infection. *Cell Microbiol* 2014;**16**:1285−300.
130. Bennett RJ, Johnson AD. Completion of a parasexual cycle in *Candida albicans* by induced chromosome loss in tetraploid strains. *EMBO J* 2003;**22**:2505−15.
131. Dvorak JA, Hall TE, Crane MSJ, Engel JC, McDaniel JP, Uriegas R. *Trypanosoma cruzi*: flow cytometric analysis. I. Analysis of total DNA/organism by means of mithramycin-induced fluorescence. *J Protozoo* 1982;**29**:430−7.
132. Messenger LA, Miles MA. Evidence and importance of genetic exchange among field populations of *Trypanosoma cruzi*. *Acta Trop* 2015;**151**:150−5.
133. Tomasini N, Diosque P. Evolution of *Trypanosoma cruzi*: clarifying hybridisations, mitochondrial introgressions and phylogenetic relationships between major lineages. *Mem Oswaldo Cruz* 2015;**110**:403−13.
134. Llewellyn MS, Lewis MD, Acosta N, Yeo M, Carrasco HJ, Segovia M, et al. *Trypanosoma cruzi* IIc: phylogenetic and phylogeographic insights from sequence and microsatellite analysis and potential impact on emergent Chagas Disease. *PLoS NTD* 2009;**3**.
135. Llewellyn MS, Miles MA, Carrasco HJ, Lewis MD, Yeo M, Vargas J, et al. Genome scale multilocus microsatellite typing of *Trypanosoma cruzi* discrete typing unit I reveals phylogeographic structure and specific genotypes linked to human infection. *PLoS Path* 2009;**5**.
136. Ocana-Mayorga S, Llewellyn MS, Costales JA, Miles MA, Grijalva MJ. Sex, subdivision, and domestic dispersal of *Trypanosoma cruzi* lineage I in Southern Ecuador. *PLoS NTD* 2010;**4**.
137. Barnabe C, Buitrago R, Bremond P, Aliaga C, Salas R, Vidaurre P, et al. Putative panmixia in restricted populations of *Trypanosoma cruzi* isolated from wild *Triatoma infestans* in Bolivia. *PLoS One* 2013;**8**.
138. Ramirez JD, Llewellyn MS. Reproductive clonality in protozoan pathogens-truth or artefact? *Mol Ecol* 2014;**23**:4195−202.
139. Ramirez JD, Llewellyn MS. Response to Tibayrenc and Ayala: reproductive clonality in protozoan pathogens − truth or artefact? *Mol Ecol* 2015;**24**:5782−4.

Downloaded From Vet-eBooks.com

140. Messenger LA, Garcia L, Vanhove M, Huaranca C, Bustamante M, Torrico M, et al. Ecological host fitting of *Trypanosoma cruzi* TcI in Bolivia: mosaic population structure, hybridization and a role for humans in Andean parasite dispersal. *Mol Ecol* 2015;**10**: 2406−22.
141. WHO. *Control of Chagas disease*. 2 ed. Geneva: World Health Organization; 2002.
142. Bargues MD, Kiisiowicz DR, Panzera F, Noireau F, Marcilla A, Perez R, et al. Origin and phylogeography of the Chagas disease main vector *Triatoma infestans* based on nuclear rDNA sequences and genome size. *Infect Genet Evol* 2006;**6**:46−62.
143. Cortez MR, Monteiro FA, Noireau F. New insights on the spread of *Triatoma infestans* from Bolivia-implications for Chagas disease emergence in the Southern Cone. *Infect Genet Evol* 2010;**10**:350−3.
144. Mauricio IL, Yeo M, Baghaei M, Doto D, Pratlong F, Zemanova E, et al. Towards multilocus sequence typing of the *Leishmania donovani* complex: resolving genotypes and haplotypes for five polymorphic metabolic enzymes (ASAT, GPI, NH1, NH2, PGD). *Int J Parasitol* 2006;**36**:757−69.
145. Lukes J, Mauricio IL, Schonian G, Dujardin JC, Soteriadou K, Dedet JP, et al. Evolutionary and geographical history of the *Leishmania donovani* complex with a revision of current taxonomy. *Proc Natl Acad Sci USA* 2007;**104**:9375−80.
146. Fraga J, Montalvo AM, De Doncker S, Dujardin JC, Van der Auwera G. Phylogeny of *Leishmania* species based on the heat-shock protein 70 gene. *Infect Genet Evol* 2010;**10**: 238−45.
147. Delgado O, Cupolillo E, BonfanteGarrido R, Silva S, Belfort E, Grimaldi G, et al. Cutaneous leishmaniasis in Venezuela caused by infection with a new hybrid between *Leishmania* (*Viannia*) *braziliensis* and *L.* (*V.*) *guyanensis*. *Mem Oswaldo Cruz* 1997;**92**: 581−2.
148. Dujardin JC, Banuls AL, Llanoscuentas A, Alvarez E, Dedoncker S, Jacquet D, et al. Putative *Leishmania* hybrids in the eastern Andean Valley of Huanuco, Peru. *Acta Trop* 1995;**59**:293−307.
149. Nolder D, Roncal N, Davies CR, Llanos-Cuentas A, Miles MA. Multiple hybrid genotypes of *Leishmania* (*Viannia*) in a focus of mucocutaneous leishmaniasis. *Am J Trop Med Hyg* 2007;**76**:573−8.
150. Evans DA, Kennedy WP, Elbihari S, Chapman CJ, Smith V, Peters W. Hybrid formation within the genus *Leishmania*? *Parasitologia* 1987;**29**:165−73.
151. Zemanova E, Jirku M, Mauricio IL, Horak A, Miles MA, Lukes J. The *Leishmania donovani* complex: genotypes of five metabolic enzymes (ICD, ME, MPI, G6PDH, and FH), new targets for multilocus sequence typing. *Int J Parasitol* 2007;**37**:149−60.
152. Schwenkenbecher JM, Wirth T, Schnur LF, Jaffe CL, Schallig H, Al-Jawabreh A, et al. Microsatellite analysis reveals genetic structure of *Leishmania tropica*. *Int J Parasitol* 2006;**36**:237−46.
153. Rougeron V, Banuls AL, Carme B, Simon S, Couppie P, Nacher M, et al. Reproductive strategies and population structure in *Leishmania*: substantial amount of sex in *Leishmania Viannia guyanensis*. *Mol Ecol* 2011;**20**:3116−27.
154. Rogers MB, Downing T, Smith BA, Imamura H, Sanders M, Svobodova M, et al. Genomic confirmation of hybridisation and recent inbreeding in a vector-isolated *Leishmania* population. *PLoS Genet* 2014;**10**:e1004092−.
155. Gelanew T, Hailu A, Schonian G, Lewis MD, Miles MA, Yeo M. Multi locus sequence and microsatellite identification of intra-specific hybrids and ancestor-like donors among natural Ethiopian isolates of *Leishmania donovani*. *Int J Parasitol* 2014;**44**: 751−7.

Downloaded From Vet-eBooks.com

156. Sadlova J, Yeo M, Seblova V, Lewis MD, Mauricio I, Volf P, et al. Visualisation of *Leishmania donovani* fluorescent hybrids during early stage development in the sand fly vector. *PLoS One* 2011;**6**.
157. Inbar E, Akopyants NS, Charmoy M, Romano A, Lawyer P, Elnaiem D-EA, et al. The mating competence of geographically diverse *Leishmania major* strains in their natural and unnatural sand fly vectors. *Plos Genet* 2013;**9**.
158. Calvo-Alvarez E, Alvarez-Velilla R, Jimenez M, Molina R, Perez-Pertejo Y, Balana-Fouce R, et al. First evidence of intraclonal genetic exchange in trypanosomatids using two Leishmania infantum fluorescent transgenic clones. *PLoS NTD* 2014;**8**.
159. Volf P, Benkova I, Myskova J, Sadlova J, Campino L, Ravel C. Increased transmission potential of *Leishmania major/Leishmania infantum* hybrids. *Int J Parasitol* 2007;**37**: 589−93.
160. Volf P, Sadlova J. Sex in *Leishmania*. *Science* 2009;**324**:1644.
161. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 2005; **309**:409−15.

Downloaded From Vet-eBooks.com

Genomic Insights Into the Past, Current, and Future Evolution of Human Parasites of the Genus *Plasmodium*

21

C.J. Sutherland [1], *S.D. Polley* [2]
[1]London School of Hygiene & Tropical Medicine, London, United Kingdom; [2]Hospital for Tropical Diseases, London, United Kingdom

1. Introduction

1.1 *Overview of* Plasmodium *Phylogeny and Description of Species Infecting* Homo sapiens

The protozoan genus *Plasmodium* comprises chromalveolate protists of the phylum Apicomplexa, order Haemosporida, family Plasmodidae. Members of the genus are obligate parasites of vertebrate hosts including lizards, snakes, birds, rodents, and primates. Amphibians, marsupials, carnivores, and ungulates are major vertebrate groups not known to host *Plasmodium* sp. parasites. Natural infections of *Homo sapiens* are caused by six species: *Plasmodium falciparum*, *Plasmodium knowlesi*, *Plasmodium malariae*, *Plasmodium vivax*, and the two closely related species *Plasmodium ovale curtisi* and *Plasmodium ovale wallikeri*, which are genetically distinct.[1,2] In naive human hosts each of these six parasites cause an acute febrile illness of varying severity and duration, known as malaria.

The evolution of the genus *Plasmodium* was punctuated by a series of host transitions, as the radiation into more than 270 current species occurred through a variety of vertebrate hosts. The primate malarias are probably the best-studied group of species, and have been well described both in natural infections of simian, ape, and human hosts,[3–6] and in experimental infections in chimpanzees, baboons, rhesus, and *Aotus* monkeys.[7] Investigations of the biochemistry and cell biology of *P. falciparum* have been possible in vitro since a system for continuous culture was devised by Trager & Jensen,[8] and this is therefore by far the best characterized of the primate malaria parasites. However, as *P. knowlesi* has now been successfully adapted to continuous culture in human erythrocytes,[9] analysis of parasite biology in vitro can be derived from two contrasting members of the genus, and is of indirect use in understanding the comparative evolutionary history of all members of the genus.

Due to the significant number of severe and fatal cases of malaria caused by *P. falciparum*, and its ability to be propagated as blood-stage parasites in vitro, Laveran's parasite remains the most closely studied of the human-infecting plasmodia.

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00021-4
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

Geographically widespread, the parasite occurs well north of the Tropic of Cancer (e.g., Afghanistan) and also south of the Tropic of Capricorn (e.g., South Africa, Namibia). Human blood-stage infections display a 48 h cycle in experimental infections in volunteers, such as those studied in detail in the Georgia State Penitentiary in the mid-20th century,[7] hence the term "malignant tertian" malaria, referring to the periodicity of fevers in *P. falciparum*. Two particular biological features of this parasite distinguish it from the other human malaria infections. The first is that parasite-encoded adhesins, capable of binding to host endothelium, are expressed on the surface of the infected erythrocyte from the mid-trophozoite stage right through to schizogony and release of merozoites. This enables the mature forms of the parasite to sequester in small blood vessels in a variety of host tissues, for a period of 30−36 h, and thus only the young trophozoites ("ring forms") are observed in smears of peripheral blood whereas, for other human malaria species, intraerythrocytic parasites at all stages of maturity circulate in the periphery. The second distinguishing feature of *P. falciparum* is the production of crescent-shaped gametocytes (transmissible stages), which do not develop in synchrony with the asexual parasite stages, but appear late in infection after an extended period of development (typically 6−10 days) sequestered in bone marrow and other endothelial beds in various tissues by virtue of the rigid undeformability of these stages, which is reversed on gaining maturity, allowing release into the peripheral circulation in order to have access to biting mosquitoes.[10] *Plasmodium vivax* infection occurs across the broadest geographical range of all the human malaria parasites, aided partly by an important survival strategy that this parasite shares with *P. ovale*: the ability of liver schizonts to arrest development and remain dormant for weeks, months, or years as hypnozoites. Reactivation of hypnozoites sometime after the primary infection following an infective mosquito bite can thus initiate a fresh blood-stage malaria infection in a subsequent season favorable for transmission to mosquitoes. This mechanism is thought particularly important for the continued transmission of *P. vivax* in temperate areas with a long winter in which anophelines are scarce or absent[3,7] or in areas with extreme or extended dry, hot seasons, such as Mauritania.[11] Malaria caused by *P. vivax* shares the 48 h periodicity of falciparum malaria, but is responsible for many fewer severe or fatal cases, hence the label "benign tertian" malaria. There is no appreciable stage-specific sequestration of parasitized erythrocytes, and all asexual and sexual forms of *P. vivax* are seen in peripheral blood smears. The lack of a reliable continuous in vitro culture system has hampered research on this parasite, but transient 48 h schizont maturation cultures have been useful in studying antimalarial drug response phenotypes in *P. vivax*.[12]

P. malariae occurs in humans throughout malaria endemic regions of the world, and has earned the name "quartan malaria" for its 72-h fever periodicity (reviewed in Collins and Jeffery[13]; Mueller et al.[14]). This species occurs at low parasite density in the peripheral blood, frequently occurs with *P. vivax* or *P. falciparum* as mixed infections, causes generally a relatively mild malaise compared to other species,[3] and is likely to be substantially underreported due to misdiagnosis as *P. vivax* or *P. falciparum*.[15] *Plasmodium malariae* also occurs as a well-described zoonosis in American monkeys, under the species designation *Plasmodium brasilianum*. Although generally considered a minor species, *P. malariae* is very common in some locations in

Downloaded From Vet-eBooks.com

PNG, Indonesia, and Africa, contributing substantially to overall malaria morbidity.[14,16,17] One of the most puzzling aspects of the biology of *P. malariae* is the ability to recur years, and even decades, after the last possible exposure of the individual to an infected mosquito bite.[18,19] Analysis of imported *P. malariae* infections in three nonendemic countries has led to the suggestion that this species may also, under certain circumstances, form latent liver-stages analogous to the hypnozoites of *P. vivax* and *P. cynomologi*.[20]

Ovale malaria is now understood to be caused by two closely related but distinct parasite species, *P. ovale curtisi* and *P. ovale wallikeri*.[1] First described by Stephens in 1922,[21] *P. ovale* spp. cause acute febrile malaria with a tertian periodicity, but only rarely are associated with severe or life-threatening complications.[14,22] Ovale malaria does not occur outside of the tropics, and is not known in the Americas, but benefits in seasonal and arid settings from the ability to form hypnozoites.[11] The absence of *P. ovale* transmission in temperate zones despite its ability to form hypnozoites suggests that the mosquito stages of this parasite are not able to complete development at lower temperatures. Relapse episodes of ovale malaria can occur months or years apart,[23,24] and therefore this species may pose a particular challenge for the eradication of malaria from sub-Saharan Africa.[1]

Dubbed by some as "the fifth human malaria," *P. knowlesi*, whose natural host is certain species of Asian macaques, is now recognized as a relatively common agent of both clinical malaria and chronic asymptomatic parasitemia in humans in Southeast Asia.[25–27] Human infections have been described in almost all countries in Southeast Asia where macaques occur.[28–32]

1.2 Population Genetics and Design of Public Health Interventions

While much attention has been paid to population-level studies of polymorphic protein-coding genes as a means to empirical identification of potential candidate vaccine antigens in both *P. falciparum* and *P. vivax*, these studies have also provided invaluable insights into the nature and complexity of historical selective forces in shaping populations of *Plasmodium* parasites.[33–37] These insights from population genetics inform not only the design of vaccines, but also the processes used to monitor the efficacy over time of vaccines put into widespread use.

Population genetic tools have also been powerfully deployed to assist understanding of the spread over recent decades of gene variants encoding parasite resistance to antimalarial drugs, particularly chloroquine (CQ) and sulfadoxine–pyrimethamine (SP).[38,39] Further, with the determination by shotgun capillary sequencing of the genome sequences of *P. falciparum*, *P. vivax*, and *P. knowlesi*,[40–42] and the recent assembly from Illumina short-read data of potential reference genome sequences for *P. ovale curtisi*, *P. ovale wallikeri*, and *P. malariae*,[2,43] genome-wide variation can now be studied among all *Plasmodium* species infecting humans. The utility of these data is enhanced by our capacity to perform direct sequencing of full parasite genomes directly from malaria patients without ex vivo culture.[44]

Downloaded From Vet-eBooks.com

1.3 Genomic Signals of Selection due to Host Immunity

Coevolution of *Plasmodium* species and their vertebrate hosts has left significant signals of selection on the host genome, well-known examples being the hemoglobin structural variants of *Homo sapiens* (such as sickle-cell anemia, the thalassemias, and G6PD deficiency), which provide a measure of protection against malaria. These genotypes have therefore become established in populations with ancient or recent history of *Plasmodium* infection risk (reviewed by Weatherall et al.[45]). Conversely, both mammalian and insect hosts have imposed selection upon the *Plasmodium* parasite genome as their respective immune systems adapt to minimize the harmful effects of parasitization. Perhaps one of the most spectacular examples of this is the genus-wide expansion of genes encoding large families of proteins involved in immune evasion. The SICAvar antigens of *P. knowlesi* were the first such proteins to be discovered, followed by a number of families in other members of the genus, notably including the PfEMP1 proteins of *P. falciparum*, encoded by the ~60-strong *var* gene family, the *surf* family ($N = 10$), and the *Pfmc-2tm* family ($N = 11$).[46] Genome analysis of a number of *Plasmodium* species genomes has identified members of the *pir* (*Plasmodium* interspersed repeat) family of small variant proteins, and among the parasites of primates these reach their greatest expansion in *P. ovale curtisi* and *P. ovale wallikeri*.[43] Thus subtelomeric gene families have expanded to encode hundreds of protein variants per genome, suggesting that the benefit of avoiding immune responses from the vertebrate host has outweighed the cost of amplifying up such large gene families, and developing the complex gene regulatory pathways required to express them effectively.

Single-copy genes throughout *Plasmodium* genomes also bear signals of immune selection, and these are exemplified by diversification of the well-characterized nucleotide sequences in both *P. falciparum* and *P. vivax* that encode the immunogenic merozoite adhesion/invasion proteins, such as AMA1, MSP2, MSP3, EBA-175, and the duffy-binding proteins.[33,35–37,45] Other single-copy genes displaying evidence of selection for diversification include those coding for the sporozoite protein thrombospondin-related protein (TRAP), and $SURFIN_{4.2}$, a protein expressed in late stage intraerythrocytic asexual parasites.[46,47] In contrast, evidence for balancing selection is not seen in population genetic analyses of the locus encoding circumsporozoite protein (CSP), despite its high degree of polymorphism.[48]

1.4 Summary of Genomic Studies of the Genus

Genomic studies per se of the genus *Plasmodium* began in the 1990s and eventually led to the full genome sequence being assembled across three sites.[40] Subsequent *Plasmodium* sp. genome projects for *P. vivax* and *P. knowlesi* utilized capillary sequencing of genomic DNA shotgun-cloned into *Escherichia coli* plasmids.[41,42] As genome sequencing technologies have become cheaper to perform at higher throughput, reduced dependence on large preparations of very pure parasite DNA, and the capacity to derive genomic information from samples in laboratory archives[49] have led to an explosion of genome-level information for *Plasmodium* spp. Adequate description of these expanding datasets and the analyses enabled by them would require an additional chapter in this book; this must wait for another opportunity.

Downloaded From Vet-eBooks.com

2. Evolution of *Plasmodium*: The Last 10 Million Years

2.1 Role of Host Transitions in Speciation Events Within the Genus

A species is a group of organism that can mate to produce fertile offspring. For speciation to occur, two populations of a species must become reproductively isolated, thereby preventing gene flow between them. The result of reproductive isolation is that over time the two populations will accumulate private mutations, which will result in the inability to form viable offspring should interpopulation breeding occur again at some time after the original separation. Allopatric speciation occurs when reproductive isolation is a direct result of the geographical separation of two (or more) populations through a physical barrier, such as an ocean. In contrast, sympatric speciation has a genetic origin, preventing admixture within two populations found in the same location. Putative mechanisms for sympatric speciation may include temporal isolation through a shift in the timing of gamete release, behavior isolation through different courtship routines, physical isolation through noncompatible genitalia, and gametic incompatibility mediated by mating incompatibility loci in plant, fungi, and marine organisms, such as the sea urchin *Echinometra* and marine diatoms.[50,51] For an obligate parasite, an obvious mechanism for sympatric speciation is host-switching. In the case of *Plasmodium*, such genetic isolation would be exacerbated if the two vertebrate hosts are targeted by different species of vector, given that genetic recombination between parasite genotypes occurs exclusively in the mosquito midgut.

Experimental infections, the comparison of *Plasmodium* and host phylogenies, and transgenic manipulation of parasites have all shed light on the role of host-switching in the evolution of the genus. Phylogenetic studies have allowed the evolution of mammalian and bird malarias to be elucidated and the common origin of these species to be calculated at between 120 and 160 million years ago (mya).[52] Such a time point is much later than the divergence of sauropsid (ancestral linage of mammals) and synapsid (ancestral lineage of birds, lizards, and snakes) vertebrate lineages some 315 million years ago, suggesting that host-switching has occurred during this time period. Human malaria species themselves appear not be monophyletic but instead have multiple independent evolutionary origins. At least three separate simian to human host transitions are apparent from the phylogenetic reconstruction of mammalian malarial mitochondrial sequences, when only *P. vivax*, *P. knowlesi*, and *P. malariae* are considered.[53] Recent studies on both bird and bat malaria species also show evidence for both host plasticity and host-switching.[54–56]

2.2 Plasmodium falciparum *and* Plasmodium reichenowi*: Divergent Host Specificities?*

The closest relative of the human malaria parasite *P. falciparum* is the chimpanzee parasite *P. reichenowi*. The ancestors of *Homo sapiens sapiens* (modern day humans) and *Pan troglodytes* (chimpanzees) are thought to have diverged around 4–8 mya according to phylogenetic and fossil evidence.[57,58] This time point has often been

Downloaded From Vet-eBooks.com

used as a date for the divergence of *P. falciparum* and *P. reichenowi*, a time point which has in turn been used to date other nodes within phylogenetic studies. When *P. reichenowi* was discovered in 1917 in the blood of wild chimpanzees and gorillas, it was morphologically indistinguishable from *P. falciparum*. However, in vitro experiments showed that humans could not be infected with *P. reichenowi*, and nor could chimpanzees be infected with *P. falciparum*—parasitized patient blood.[59] The two infections were therefore deemed to be separate species, a view that was subsequently supported by genetic analysis of the *csp* (circumsporozoite protein) locus.[60]

As early as 1977, differences in the erythrocyte receptors of humans and chimpanzees were postulated as the driving force behind the apparent species specificity of *P. reichenowi* and *P. falciparum*,[61] and this has since been related to differences in the sialic acid structure of Glycophorin A between humans and chimpanzees.[59] However, recent insights gained from studies of the genomes of the Laveranian parasites *P. reichenowi* and *P. gaboni* from DNA fragments isolated from wild ape feces show that *P. falciparum* has only recently transferred into *Homo sapiens*, and shows relatively little genetic diversity compared to the former two species, or even compared to *P. falciparum*—like parasites from wild gorillas.[62,63] This strongly suggests a more profound basis to speciation than Miller's concept based on specialized erythrocyte invasion, which is more likely to be a recent adaptation after the original host transition.

While early experiments showed that chimpanzees were refractory to *P. falciparum*, experimental infections can be established in splenetcomized chimpanzees. Indeed, several isolates of *P. falciparum* including Vietnam Oak Knoll and Uganda Palo Alto[64,65] have been adapted by serial passage through New World monkey hosts for use with animal models. The last few years have also seen the discovery of *P. knowlesi* infections in humans,[25] confirmation of *P. falciparum* infection in both great apes and monkeys,[4,62] and both *P. malariae* and *P. vivax* in chimpanzees,[66,67] suggesting that the host specificity of certain *Plasmodium* species may be more fluid than was once thought: such infections, at the very least, suggest that mechanisms such as specific adaptation to erythrocytic receptors specific to a particular primate host may be insufficient to fully explain host-switching and potential sympatric speciation events.

What alternatives are there to host-switching that would facilitate the genetic isolation of sympatric populations? Mutations within mating incompatibility loci are known to produce genetic isolation within sympatric populations of animals and plants. The surface gamete proteins Pfs48/45 are known to be critical for the fertility of male microgametes in *P. falciparum*,[68] and gene distributions at this locus show significant skewing at the population level.[69] Following fertilization, a diploid ookinete is formed, which buries into the mosquito midgut and undergoes meiotic reduction to form a sporozoite-filled oocyst. Analysis of Pfs48/45 allele frequencies in oocysts gathered from Tanzanian mosquitos shows a significantly higher inbreeding coefficients (F(IS)) compared with 11 unlinked microsatellite loci locus. Thus it would appear that assortative mating is occurring in these populations,[70] raising the possibility that genetic isolation may occur through mutations in gametic surface receptors at the time of host-switching by the novel species.

Downloaded From Vet-eBooks.com

2.3 *Speciation Between* Plasmodium ovale curtisi *and* Plasmodium ovale wallikeri*: Separate Host Transitions?*

The parasites previously defined as *P. ovale*, on the basis of their morphology when viewed on Giemsa-stained blood films, are now known to comprise two genetically distinct, nonrecombining species that diverged around 2 mya. These two newly recognized species were shown to be sympatric at country level in their distribution, thus ruling out classic allopatric barriers as an explanation for the current pattern of endemicity, and suggesting that reproductive isolation between the two parasites is due to other mechanisms, such as host-partitioning.[1] Nevertheless, as only country-level distribution data was available for this first study of the dimorphic parasite species, it could not be ruled out that continuing physical separation between the two species is maintained due to more subtle mechanisms, such as microgeographical discontinuities or specific habitat or seasonal requirements preventing admixture of the two types in the present. Subsequent examination of parasites from clinics in Congo-Brazzaville, and community parasitological surveys in Equatorial Guinea and two sites in Uganda demonstrated conclusively that both *P. ovale curtisi* and *P. ovale wallikeri* occur in different people in the same village at the same time, providing stronger support for the existence of a robust biological barrier (or barriers) between the two parasite species.[71]

It is prudent to recognize that whatever the current barriers to mating and recombination might be, it is not necessarily true that these were also the *cause* of the original speciation event. In our view, the most parsimonious explanation for the existence of these two related but distinct parasite lineages in the same primate host, *Homo*, is that their common ancestor in the most recent nonhuman primate host (*P. ovale sensu stricto*) underwent successful transit to *Homo* on two separate occasions. The two lineages antecedent to the *P. ovale curtisi* and *P. ovale wallikeri* species were thus prevented from recombination because each occupied a different primate host, and so the two lineages never occurred simultaneously in a single mosquito blood-meal, the absolute requirement for hybridization. During this time apart, differences arose, which meant that recombination was no longer possible, once both lineages had become parasites of humans. It is also valid to argue an alternative to our favored "two transition hypothesis," namely that separation was due to a geographical separation of two populations *after* host transition to *Homo* (or the ancestors of *Homo*). This could be dubbed the "transient allopatry hypothesis," in that it requires sympatric, early *Homo*-parasitizing (hypothetical) *P. ovale s.s.* populations to have developed into two geographically isolated lineages that came back into sympatry due to migration of the hominid hosts or expansion of the separated parasite populations. During the period of separation, the pair of lineages underwent sufficient divergent evolution for differences to arise, and these now keep the two lineages apart, despite their current close cohabitation across a global range.[1,2,71,72]

What could be the nature of these differences that maintain the species barrier between *P. ovale curtisi* and *P. ovale wallikeri* now that they exist in sympatry? This is a matter of great interest, and a number of explanations can be hypothesized, and eventually tested. Here, we briefly consider five (see also Sutherland,[2] Box 1).

Downloaded From Vet-eBooks.com

- The two species remain physically separated by classic allopatry, or by a cryptic allopatry brought about by different patterns of seasonal emergence.
- Special requirement of each species for different species of mosquito vectors.
- Special requirement of each species for different human subsets defined by blood groups of erythrocyte surface molecules.
- *P. ovale curtisi* and *P. ovale wallikeri* essentially propagate clonally due to very frequent self-fertilization, and this explains why both exhibit very little intraspecies polymorphism.
- Through genetic drift, divergence of specific molecules required during the mating process, or gross changes at chromosomal level that prevent viable chromosome pairing at meiosis.

The last of these five hypotheses is the only one we now consider plausible: that divergence in genome primary and secondary structures during a period of separation prevents effective meiotic pairing between the two current lineages, now that they are circulating in the same host species (*Homo sapiens*). Thus recombination does not occur between *P. ovale curtisi* and *P. ovale wallikeri*. We have previously postulated that human blood groups may delineate mutually exclusive subsets of human hosts.[1] Species-specific restriction to different subsets of the human population had appeared to be a plausible mechanism for the current maintenance of the species barrier between *P. ovale curtisi* and *P. ovale wallikeri*. A precedent exists in the form of the well-described reliance of *P. vivax* on the Duffy blood group antigen for invasion of host reticulocytes. Weak support for this notion came from the observation that the two ovale homologs of the *P. vivax* gene encoding the reticulocyte-binding protein *pvrb2*, implicated in blood-stage invasion and thus potentially host cell selection, have accumulated a number of nonsynonymous mutations.[71] Thus *P. ovale curtisi* and *P. ovale wallikeri* might be expected to exhibit phenotypic differences in erythrocyte/reticulocyte invasion, including different patterns of blood group restriction in selecting host red blood cells for invasion.[1] However, evidence from field studies shows that both species can be identified simultaneously in a single individual,[17,73] findings that disprove the blood group restriction hypothesis.

Species-specific restriction of *P. ovale curtisi* and *P. ovale wallikeri* to different species of *Anopheles* mosquito is also unable to explain the lack of recombination between the two parasites. There are a large variety of competent malaria vector species within the genus *Anopheles*, and the malaria vectors of medical importance differ within and between nations, and within and between continents. Thus the distribution of both *P. ovale curtisi* and *P. ovale wallikeri* in Asia, Africa, and the Pacific demonstrates unequivocally that both parasites are transmitted by a variety of mosquito species across this broad range.[1,32,73] Mosquito host restriction, although possibly important in some endemic localities, therefore cannot explain the observed lack of recombination between the two *ovale* species.

Allopatric speciation events can generate two related taxa, physically separated, which can become secondarily sympatric due to migration or changes in the extent of suitable habitat (Section 2.1). If during the period of separation, substantial genetic drift occurs in the sequences of genes determining mating compatibility, or chromosomal rearrangements occur, interrupting synteny such that hybrid offspring are extremely unlikely to be viable, then the species barrier will remain intact, and recombination between the two forms will not occur. As argued earlier, we consider that the

Downloaded From Vet-eBooks.com

most likely reason for the evolution of two forms of ovale malaria parasite is that at least two independent host transitions into ancestors of modern humans occurred, separated by a lengthy period of time. This scenario is of course analogous to allopatry, in that the two lineages would have been "physically" separated by occupying different hosts during the period between the first and second transitions. Thus the lack of recombination between 21st century human-dwelling populations of *P. ovale curtisi* and *P. ovale wallikeri* is most likely due to genomic changes accrued over this period of separation, thus rendering meiotic chromosome pairing impossible, and hybrid zygotes unviable. The recent genomic data of Ansari et al.[43] does identify divergence among the greatly expanded subtelomeric *oir* genes, encoding members of the PIR super family of variant antigens found in all members of the genus, and areas of syntenic discontinuity, that together could hinder effective homologous chromosome pairing in heterologous matings between the two ovale parasite species.

These genomic studies are therefore poised to answer important questions about genome-wide interspecies polymorphisms—which loci have diverged the most between *P. ovale curtisi* and *P. ovale wallikeri*? Are erythrocyte invasion molecules prominent among them? Are the greatly expanded *oir* loci, located in the subtelomeres that accrue diversity and structural rearrangements at a faster rate than other chromosomal regions, helping to prevent efficient meiotic pairing between the two species?[2] Finally, we will also be able to compare contemporary 21st century isolates of *P. ovale curtisi* with the partial capillary−sequenced isolate collected in Nigeria over 30 years previously (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/p_ovale), and thus also gain new understanding of intraspecies polymorphism, in both time and space, at the genome level. Close examination of intraspecies polymorphism will directly address the hypothesis of clonal propagation in ovale parasites. It is certainly true that, to date, there is little evidence that either ovale parasite species is highly diverse genetically, and our first study of interspecies genetic variation found only a very low level of diversity at the seven loci examined.[1] However, the genes studied were mostly identified due to their high degree of sequence conservation with homologous loci in other better characterized *Plasmodium* species; such genes are likely to be under purifying selection and thus poor candidates with which to estimate levels of linkage disequilibrium, and coefficients of inbreeding. Some intraspecific polymorphism has now been found in the *potra* locus,[71,74] *cox1*[73] and *msp1*.[75] Thus, although it remains possible that frequent clonal propagation has exacerbated the isolation of these two genomes, this is very unlikely to be an explanation for the existence of a species barrier per se.

2.4 Importance of Host Specificity

Understanding the mechanisms of host specificity, and host-switching, will allow a greater understanding of the pathology of *Plasmodium* spp. infections, identifying key genes required for both asexual and sexual reproduction, and possible targets for disruption through the production of novel vaccines and chemotherapeutics. The establishment of SARS, HIV, Ebola, and Zika viral infections in human populations shows how host-switching can lead to the emergence of a highly virulent pathogen, and a relatively recent switch to human hosts might explain the increased

Downloaded From Vet-eBooks.com

pathogenicity of *P. falciparum* compared to *P. ovale*, *P. vivax*, and *P. malariae*. Ultimately, detailed understanding of the parasite genes contributing to the vertebrate host specificity of each *Plasmodium* parasite species may come from careful comparisons among the genomes now being sequenced. Given the current understanding of the importance of erythrocyte-binding proteins and their specificity for certain host receptors (Section 2.2), elucidation of the reticulocyte-binding proteins, Duffy-binding proteins, and erythrocyte-binding protein families in each species is a good place to start. In fact, it now appears that cultured lines of *P. knowlesi* that have been adapted to propagate in human erythrocytes express particular variants of the reticulocyte-binding protein, suggesting different phenotypes are favored compared to parasites adapted to macaque erythrocytes.[9,76,77] However, we should bear in mind that an uncomfortably large proportion of each *Plasmodium* genome thus far sequenced encodes for "hypothetical proteins" of unknown function; it may be that comparative empirical approaches are required to identify many of the key molecules involved in parasite host specificity.[78]

3. Evolution of *Plasmodium*: The 21st Century in Three Courses

3.1 Entree: Zoonoses—A Legacy of Habitat Destruction for Wild Primates?

Host transitions in malaria parasites require contact between the novel vertebrate host species and insect vectors that have bitten infected individuals of the primary vertebrate host species. Thus human (and prehuman) migration, settlement, and the resulting encroachment of human activity into the habitats of different nonhuman primates have been the probable driving force behind the prehistorical transitions discussed in Section 2. In the 21st century, human migrations still occur, and encroachment into the habitats of wild simian and ape populations continue. Coupled with a recent improvement in our ability to discern the presence of unusual *Plasmodium* species in both humans and nonhuman primates, these continued close encounters between humans, and the parasites of beasts, will be more commonly recognized. Further, the destruction of habitat and the resulting decline in numbers of the great apes means that the parasite species dependent on them as hosts are under selective pressure to expand their host range. This may increase the likelihood of new transitions into human-centered parasitism.

On the other hand, the human parasite *P. falciparum* is known to be present in wild African apes, with identification of this "human" parasite in bonobos (*Pan paniscus*) from the Democratic Republic of Congo,[79] in *Gorilla gorilla diehli* from Cameroon and *G. g. gorilla* from Gabon[6] and other central African locations.[62] Genetic analyses of these strongly suggest continued cycling of the Laverania grouping of malaria parasites among different hominids, including *Homo*, in Africa. Conversely, as these studies also identified additional examples of *P. reichenowi* in three subspecies of *Pan* in Ivory Coast and Cameroon, and a closely related parasite in a bonobo from

Downloaded From Vet-eBooks.com

DRC, as well as two new species, *P. billbrayi* and *P. billcollinsi*, the possibility now occurs that a variety of hitherto overlooked parasite species, with *Pan* and *Gorilla* as primary hosts, may be zoonotic in humans. This poses two interesting questions. First, have we seriously underestimated the number of malaria parasite species that naturally infect humans in central Africa, by ignoring the possibility of frequent zoonotic infections? Secondly, will the great apes provide an eradication-proof reservoir of human malaria parasites, particularly *P. falciparum*, *P. ovale curtisi*, and *P. ovale wallikeri* (Section 4; Duval et al.[4,5])?

The recent descriptions of contemporary wild isolates of *P. reichenowi*[6,63] raises the possibility that human infections with this parasite may occasionally occur in some parts of Africa, but have failed to be recognized. This is not least because remnant populations of the nonhuman hominids exist in relatively remote locations, where people suffering from malaria who do come in contact with health services are likely to be treated (if at all) without any diagnosis, and certainly without a species-specific one.[1] The prospect of such previously unrecognized zoonoses occurring is an interesting one from a scientific point of view, but may also have important evolutionary and public health consequences. Although at least one recent study found no evidence of ape parasites among human communities near populations of wild apes in Gabon,[80] studies of likely vectors in the forest canopy do suggest there is no entomological barrier to zoonotic transmission of these ape parasite species to humans.[81] As long as such infections are at least possible, the evolutionary change to permit efficient human-to-human transmission of a previously ape-confined species may occur. From a public health point of view, this could lead to rapid expansion of a "new" malaria pathogen in human populations with no species-specific immunity; this threat is even more relevant were malaria elimination/eradication to occur in these regions in the near future. Removal of human parasite species through vector control, immunization, and effective drug deployment will lead to a human population with no naturally acquired immunity, and perhaps with greatly reduced access to effective malaria treatment as the number of cases dwindle. Thus understanding of these zoonotic pathogens should be vigorously pursued not only because of its great scientific interest, which would be enough for most of us, but through human self-interest, which may be needed to persuade science funders.

3.1.1 *Emergence of* Plasmodium knowlesi *Zoonosis*

As mentioned earlier (Section 1.1), the story of *P. knowlesi* in Southeast Asia is one particularly important example, becoming well understood only during 2000s, of a zoonotic malaria parasite being recognized as a significant public health problem for humans. The parasite was thought to only naturally infect long-tailed and pig-tailed macaques, which are widely distributed across Southeast Asia and in which infection is benign,[7] despite one or two sporadic reports of human cases in Singapore and peninsular Malaysia.[3] It had been demonstrated during the use of induced malaria infections for syphilis therapy in the early 20th century that *P. knowlesi* would infect humans with effective blood-stage multiplication, and cause full-blown clinical malaria,[3] but the true extent of naturally occurring human cases had not been understood, and

Downloaded From Vet-eBooks.com

was assumed to be negligible. Molecular investigations during a malaria epidemic in Malaysian Borneo in the late 1990s confirmed *P. knowlesi* as the causative agent of a significant number of human cases.[25] Molecular studies have since identified human cases of *P. knowlesi* in several Asian countries in addition to Malaysia, including Myanmar, the Philippines, Singapore, Vietnam, and Thailand.[28–31,82] *P. knowlesi* infections in humans can cause severe and fatal disease,[26] but it is now becoming clear from studies in Malaysian Borneo,[27] Myanmar, and Sumatra (Lubis and Sutherland, unpublished results) that *P. knowlesi* infections can occur widely as asymptomatic chronic infections in communities in proximity to macaque populations.

It is probably the case that infections of humans in areas where macaques abound have always occurred, and that human *P. knowlesi* infections may have been either ignored, or extensively misdiagnosed in the past.[83] However, as there is as yet no evidence of human-to-mosquito transmission, the disease is best described as a well-established zoonosis with a measurable public health impact in some areas. Should efficient human-to-vector transmission arise de novo, then *P. knowlesi* could easily spread via human movements beyond the current range, which for now is absolutely restricted by the distribution of suitable host macaque species. Such a "break-out" of this species may carry a threat of substantial mortality, as this parasite has a very rapid replication cycle and can reach life-threatening parasite densities in a small proportion of infected human individuals.[26]

3.2 Plat du Jour: Chemotherapy and the Evolution of Drug-Resistant Parasites

3.2.1 Lessons Learned From the Evolution of Parasite Resistance to Chloroquine

A well-studied example of recent evolution in the genus *Plasmodium* is the worldwide development of resistance to the antimalarial drug CQ during the latter half of the 20th century. The evolutionary aspects of this drug selection has been well examined,[84] but some important principles can be drawn out, which are relevant to understanding the likely impact of antimalarial drugs in the present century.

First, although there is evidence that CQ-resistant alleles of the transporter gene *pfcrt* evolved in different lineages of *P. falciparum* on multiple occasions, one or two of these alleles were particularly successful, spreading through contiguous parasite populations, and moving intercontinental distances, presumably in human travelers, to dominate parasite populations worldwide.[85] Interestingly, in the case of the allele which encodes the amino acid haplotype SVMNT at codons 72–76 of *pfcrt*, it now seems likely that heavy use of the related amino-quinoline drug amodiaquine (AQ) in some countries in the mid-20th century provides a plausible alternative explanation for the spread of this parasite type.[86] Thus CQ resistance per se may be a secondary characteristic of the CRT protein variant encoded by this allele.

Secondly, antimalarial selection is far from uniform in endemic areas, and so there are both temporal and geographical variations in the intensity of drug pressure, broadly meaning the proportion of infected people likely to use a particular drug type. This

Downloaded From Vet-eBooks.com

modulates the selective advantage enjoyed by parasites carrying drug-resistant alleles of key genes, particularly if there is a fitness cost to the development of resistance. Thus poor access to treatment, high levels of antiparasite immunity leading to asymptomatic parasite carriage (particularly in African adults), and extended periods in the dry season with no transmission of new infections have contributed to the successful survival and continual transmission of wild-type, CQ-sensitive *P. falciparum* in many parts of Africa throughout the period that CQ was the main treatment option for malaria.[87] In the extreme case of complete removal of CQ from an entire health system, as was achieved in Malawi, the fitness advantage of wild-type CQ-sensitive parasites ensures their relatively rapid resurgence to replace CQ-resistant genotypes that no longer have a survival advantage.[88]

Thirdly, different antimalarial regimens generate different, sometimes opposite, selective forces on the parasite genome.[89,90] Thus as drug use diversifies in the current post-CQ era, the parasite genome appears to be showing evidence of "balancing selection" at drug resistance loci, a concept in direct opposition to the theoretical notion that drug pressure always drives advantageous alleles to "fixation" (Section 3.2.2).

Finally, although genetically determined resistance to CQ has been described in several species of *Plasmodium*, only in *P. falciparum* are mutations in the *crt* locus directly linked to altered response to the drug. Stable resistance to CQ in both *P. vivax* and in *P. chabaudi* has been described, but the *pvcrt* and *pccrt* loci, respectively, remain unchanged in resistant parasites.[91,92] This is in contrast to resistance to the antifolate drugs pyrimethamine and sulfadoxine, which involve nonsynonymous mutations in the *dhfr* and *dhps* loci of each *Plasmodium* species so far investigated. Thus the use of model host−parasite systems, and even in vitro studies of drug sensitivity, may not provide adequate understanding of drug resistance of *Plasmodium* parasites in vivo.

3.2.2 Evolution of Parasite Resistance in the Artemisinin Combination Therapy Era

As CQ use across malaria endemic regions becomes less common, there is in effect a rapid diversification of the selective drug pressure upon malaria parasites, after decades of intense directional selection on the *Plasmodium* genomes for resistance to CQ and to the antifolate-fixed combination drug SP. A major contributor to this diversification is the introduction of ACT by a majority of governments as the main replacement for CQ and SP in the formal health service of most endemic countries. The most frequently used ACT in Africa are artemether−lumefantrine (AL) and artesunate−amodiaquine (ASAQ), and, importantly, there is good evidence that the partner drugs lumefantrine and amodiaquine exert strong selection on the *pfmdr1* locus in opposite directions. This gene encodes the ABC transporter PgH1, which is homologous to the multidrug-resistant proteins of mammalian cells, and involved in modulating the effects of multiple antimalarial drugs. Thus whereas recurrence of parasitemia after AL treatment is associated with the *pfmdr1* haplotype NFD at codons 86, 184, and 1246, amodiaquine selects for the haplotype YYY (Table 4 in Humphreys et al.[90]). At the same time, in the private sector, both artemisinin monotherapies and a

Downloaded From Vet-eBooks.com

variety of ACT are available in shops and pharmacies, many of whom continue to sell CQ, SP, amodiaquine, oral quinine, and other antimalarials. Newer ACT formulations, including dihydroartemisinin−piperaquine and artesunate−pyronaridine have also entered the marketplace, and have been adopted by Government health services. The selective pressure exerted by these regimens on *pfmdr1*, *pfcrt*, and other loci of importance in determining drug response in *P. falciparum* is now starting to become clear. Studies from the Greater Mekong area and in East Africa provide evidence that partner drugs piperaquine, amodiaquine, and lumefantrine are able to exert pressure on populations of *P. falciparum*,[93–95] leading to change in population prevalence estimates of key resistance-associated genoypes.[96]

The previous paragraph gives consideration to the variety of partner drugs used in different ACT formulations. However, what about the selective effect of the use of the artemisinin compounds themselves? The emergence of *P. falciparum* parasites in Cambodia with markedly increased survival time in vivo following artesunate monotherapy has caused great alarm, and there have been important efforts to systematically monitor parasite clearance times across the global range of *P. falciparum*, to both artesunate alone and to ACT.[97,98] Despite the coverage afforded by partner drugs, the almost universal adoption of ACT in endemic countries does effectively guarantee that selection on parasite genomes by artemisinins is essentially worldwide. Interestingly, these slow clearance phenotypes, associated with specific mutations in the propeller domain of the *P. falciparum* K13 kelch protein, appear to have arisen more than once in Southeast Asia under artemisinin selection, rather than spreading from a single focus.[99] Further, in African settings, no association between treatment failures and *pfk13* mutations have been shown to date.[100] Rather, multigenic haplotypes including variants at the loci *pfcrt*, *pfmdr1*, *pfap2mu*, and *pfubp1* have been implicated in at least one East African study.[94] Thus, although *pfk13* polymorphisms are not currently a threat to African parasite drug susceptibility, it remains possible that parasites with extended post-artemisinin survival times will continue to evolve, and could spread rapidly over intercontinental distances, as was seen for CQ resistance in the 20th century. Several strategies could minimize the impact of such newly evolving ART resistance, and these largely rely on the expectation that evolution of resistance involves a fitness cost for the parasite:

- Partner drug diversification should be deployed in ACT, so that parasites less susceptible to one combination are likely to be contained if treated with a different combination.
- The use of triple combination extended regimens should be considered where indicated by falling ACT efficacy, deploying two different ACT sequentially, so that treatment comprises 6 days of artemisinin and two partner drugs in total.
- Continued development of new drugs must take place, for additional partners to combine with the artemisinins, and for potent new compounds to replace the artemisinins should their failure be catastrophic at some point in the future.

Following these three strategies will assist in managing the evolution of resistance in malaria parasites, in order to maximize the duration of the public health benefit currently enjoyed due to the effectiveness of ACT.

Downloaded From Vet-eBooks.com

3.3 Dessert: Selection for Immunological Escape Variants by Malaria Vaccines: A Real or Imagined Threat?

At the time of writing, 2016, a single registered vaccine, RTS, S/AS02, which offers partial protection against clinical malaria episodes in young children, is due to become available. However, the increased investment in malaria vaccine development from private, governmental, and intergovernmental funding agencies since 2000, which led to the successful development of this vaccine, has pushed malaria vaccine research forward and greatly improved the chances of having other products to deploy during the next decade. A theoretical obstacle to effective vaccination against malaria is that the parasite will exploit the existing polymorphism of target antigens against which vaccines are currently being developed to escape vaccine-elicited immunity, and there is concern that this problem may not be easily overcome.[101,102] For the RTS, S vaccine, there has been no evidence in any study thus far of strain selection for parasites encoding particular escape variants of the CSP vaccine antigen, but there is evidence that the vaccine is more effective against parasites carrying *pfcsp* genes similar or identical in sequence to that in the vaccine construct.[103] Thus, allele selection may occur in the future if the vaccine is widely deployed. An advantage of whole parasite vaccine strategies, such as the SPZ series of sporozoite vaccines currently under development, is that such genotype selection is far less likely due to the complexity of the antigens delivered during vaccination.[104]

4. Evolution of *Plasmodium*, and the Eradication Agenda

As malaria eradication programs are envisaged for implementation around the malaria endemic world, evolutionary biologists are pondering the potential of uniform, large-scale interventions to force the parasite genome through selection bottlenecks. An example of such a bottleneck is the impact of the antimalarial CQ, which for decades in the 20th century exerted monolithic selection on the *P. falciparum* genome for a handful of advantageous alleles of *pfcrt* and *pfmdr1*. Some questions arise:

Will the eradication agenda pose a straightforward challenge that the parasite is more than able to meet? To avoid this outcome, it is essential that multiple tools are deployed, including not just antiparasite but also antivector methods, and interventions to prevent human behaviors that enhance risk of exposure to malaria.

Will malaria continue to exert selection for new mutations on the human genome, as it has in the past? This seems unlikely, as relatively few people are prevented from reproduction by malaria, and hopes are high that efforts toward eradication will rapidly bring this number down even further. Nevertheless, there is ample evidence that already existing variants, such as the thalassemias, G6PD deficiency, and sickle-cell Hb, are maintained at stable frequencies in human populations by offering protection against malaria.

Will zoonotic primate malaria infections lead to a reservoir of hit-and-run malaria cases on the fringes of forests in Africa and Asia, and might these species adapt to anthropocentric transmission? This fascinating and potentially dangerous scenario

Downloaded From Vet-eBooks.com

cannot be ruled out, and, should human malaria be eradicated at some time in the future, will require careful vigilance in human populations that frequent the forest fringes of Southeast Asia and central Africa.

References

1. Sutherland CJ, Tanomsing N, Nolder D, Oguike M, Jennison C, Pukrittayakamee S, et al. Two non-recombining sympatric forms of the human malaria parasite *Plasmodium ovale* occur globally. *J Infect Dis* 2010;**201**:1544−50.
2. Sutherland CJ. Persistent parasitism: the adaptive biology of malariae and ovale malaria. *Trends Parasitol* 2016 [in press].
3. Garnham PCC. *Malaria parasites and other haemosporidia*. Oxford: Blackwell; 1966.
4. Duval L, Nerrienet E, Rousset D, Sadeuh Mba SA, Houze S, Fourment M, et al. Chimpanzee malaria parasites related to *Plasmodium ovale* in Africa. *PLoS One* 2009;**4**: e5520.
5. Duval L, Fourment M, Nerrienet E, Rousset D, Sadeuh SA, Goodman SM, et al. African apes as reservoirs of *Plasmodium falciparum* and the origin and diversification of the *Laverania* subgenus. *Proc Natl Acad Sci USA* 2010;**107**:10561−6.
6. Prugnolle F, Durand P, Neel C, Ollomo B, Ayala FJ, Arnathau C, et al. African great apes are natural hosts of multiple related malaria species, including *Plasmodium falciparum*. *Proc Natl Acad Sci USA* 2010;**107**:1458−63.
7. Coatney GR, Collins WE, Warren M, Contacos PG. *The primate malarias Division of Parasitic Disease, producers. Version 1.0*. Atlanta, GA, USA: CDC; 2003. CD-ROM. [Originally published 1971].
8. Trager W, Jensen JB. Human malaria parasites in continuous culture. *Science* 1976;**193**: 673−5.
9. Moon RW, Hall J, Rangkuti F, Ho YS, Almond N, Mitchell GH, et al. Adaptation of the genetically tractable malaria pathogen *Plasmodium knowlesi* to continuous culture in human erythrocytes. *Proc Natl Acad Sci USA* 2013;**110**:531−6.
10. Tibúrcio M, Silvestrini F, Bertuccini L, Sander AF, Turner L, Lavstsen T, et al. Early gametocytes of the malaria parasite *Plasmodium falciparum* specifically remodel the adhesive properties of infected erythrocyte surface. *Cell Microbiol* 2013; **15**:647−59.
11. Lekweiry KM, Abdallahi MO, Ba H, Arnathau C, Durand P, Trape JF, et al. Preliminary study of malaria incidence in Nouakchott, Mauritania. *Malar J* 2009;**8**:92.
12. Wirjanata G, Handayuni I, Prayoga P, Apriyanti D, Chalfein F, Sebayang BF, et al. Quantification of *Plasmodium* ex vivo drug susceptibility by flow cytometry. *Malar J* 2015;**14**:417.
13. Collins WE, Jeffery GM. *Plasmodium malariae*: parasite and disease. *Clin Microbiol Rev* 2007;**20**:579−92.
14. Mueller I, Zimmerman PA, Reeder JC. *Plasmodium malariae* and *Plasmodium ovale* − the 'bashful' malaria parasites. *Trends Parasitol* 2007;**23**:278−83.
15. Lindo JF, Bryce JH, Ducasse MB, Howitt C, Barrett DM, Morales JL, et al. *Plasmodium malariae* in Haitian refugees, Jamaica. *Emerg Infect Dis* 2007;**13**:931−3.
16. Bruce MC, Macheso A, Kelly-Hope LA, Nkhoma S, McConnachie A, Molyneux ME. Effect of transmission setting and mixed species infections on clinical measures of malaria in Malawi. *PLoS One* 2008;**3**:e2775.

Downloaded From Vet-eBooks.com

17. Dinko B, Oguike MC, Larbi JB, Bousema JT, Sutherland CJ. Persistent detection of *Plasmodium falciparum*, *P. malariae*, *P. ovale curtisi* and *P. ovale wallikeri* after ACT treatment of asymptomatic Ghanaian school-children. *Int J Parasitol DDR* 2013;**3**:45−50.
18. Tsuchida H, Yamaguchi K, Yamamoto S, Ebisawa I. Quartan malaria following splenectomy 36 years after infection. *Am J Trop Med Hyg* 1982;**31**:163−5.
19. Vinetz JM, Li J, McCutchan TF, Kaslow DC. *Plasmodium malariae* infection in an asymptomatic 74-year-old Greek woman with splenomegaly. *N Engl J Med* 1998;**338**: 367−71.
20. Teo BH, Lansdell P, Smith V, Blaze M, Nolder D, Beshir KB, et al. Delayed onset of symptoms and atovaquone-proguanil chemoprophylaxis breakthrough by *Plasmodium malariae* in the absence of mutation at codon 268 of *pmcytb*. *PLoS Negl Trop Dis* 2015;**9**: e0004068.
21. Stephens JWW. A new malaria parasite of man. *Ann Trop Med Parasitol* 1922;**16**:383−6.
22. Collins WE, Jeffery GM. *Plasmodium ovale*: parasite and disease. *Clin Microbiol Rev* 2005;**18**:570−81.
23. Coldren RL, Jongsakul K, Vayakornvichit S, Noedl H, Fukuda MM. Apparent relapse of imported *Plasmodium ovale* malaria in a pregnant woman. *Am J Trop Med Hyg* 2007;**77**: 992−4.
24. Nolder D, Oguike MC, Maxwell-Scott H, Niyazi HA, Smith V, Chiodini PL, et al. An observational study of malaria in British travellers: *Plasmodium ovale wallikeri* and *Plasmodium ovale curtisi* differ significantly in the duration of latency. *BMJ Open* 2014;**3**: e002711.
25. Singh B, Kim Sung L, Matusop A, Radhakrishnan A, Shamsul SS, Cox-Singh J, et al. A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. *Lancet* 2004;**363**:1017−24.
26. Cox-Singh J, Davis TM, Lee KS, Shamsul SS, Matusop A, Ratnam S, et al. *Plasmodium knowlesi* malaria in humans is widely distributed and potentially life threatening. *Clin Infect Dis* 2008;**46**:165−71.
27. Fornace KM, Nuin NA, Betson M, Grigg MJ, William T, Anstey NM, et al. Asymptomatic and submicroscopic carriage of *Plasmodium knowlesi* malaria in household and community members of clinical cases in Sabah, Malaysia. *J Infect Dis* 2016;**213**:784−7.
28. Luchavez J, Espino F, Curameng P, Espina R, Bell D, Chiodini P, et al. Human infections with *Plasmodium knowlesi*, the Philippines. *Emerg Infect Dis* 2008;**14**:811−3.
29. Ng OT, Ooi EE, Lee CC, Lee PJ, Ng LC, Pei SW, et al. Naturally acquired human *Plasmodium knowlesi* infection, Singapore. *Emerg Infect Dis* 2008;**14**:814−6.
30. Putaporntip C, Hongsrimuang T, Seethamchai S, Kobasa T, Limkittikul K, Cui L, et al. Differential prevalence of *Plasmodium* infections and cryptic *Plasmodium knowlesi* malaria in humans in Thailand. *J Infect Dis* 2009;**199**:1143−50.
31. Van den Eede P, Van HN, Van Overmeir C, Vythilingam I, Duc TN, Hung le X, et al. Human *Plasmodium knowlesi* infections in young children in central Vietnam. *Malar J* 2009;**8**:249.
32. Setiadi W, Sudoyo H, Trimarsanto H, Sihite BA, Saragih RJ, Juliawaty R, et al. A zoonotic human infection with simian malaria, *Plasmodium knowlesi*, in Central Kalimantan, Indonesia. *Malar J* 2016;**15**:218.
33. Polley SD, Conway DJ. Strong diversifying selection on domains of the *Plasmodium falciparum* apical membrane antigen 1 gene. *Genetics* 2001;**158**:1505−12.
34. Rayner JC, Corredor V, Feldman D, Ingravallo P, Iderabdullah F, Galinski MR, et al. Extensive polymorphism in the *Plasmodium vivax* merozoite surface coat protein MSP-3α is limited to specific domains. *Parasitology* 2002;**125**:393−405.

Downloaded From Vet-eBooks.com

35. Ord R, Polley S, Tami A, Sutherland C. High sequence diversity and evidence of balancing selection in the *Pvmsp3α* gene of *Plasmodium vivax* in the Venezuelan Amazon. *Mol Biochem Parasitol* 2005;**144**:89–93.
36. Polley SD, Tetteh KK, Lloyd JM, Akpogheneta OJ, Greenwood BM, Bojang KA, et al. *Plasmodium falciparum* merozoite surface protein 3 is a target of allele-specific immunity and alleles are maintained by natural selection. *J Infect Dis* 2007;**195**:279–87.
37. Ord RL, Tami A, Sutherland CJ. *ama1* genes of sympatric *Plasmodium vivax* and *P. falciparum* from Venezuela differ significantly in genetic diversity and recombination frequency. *PLoS One* 2008;**3**:e3366.
38. Roper C, Pearce R, Nair S, Sharp B, Nosten F, Anderson T. Intercontinental spread of pyrimethamine-resistant malaria. *Science* 2004;**305**:1124.
39. Anderson TJ, Roper C. The origins and spread of antimalarial drug resistance: lessons for policy makers. *Acta Trop* 2005;**94**:269–70.
40. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 2002;**419**:498–511.
41. Carlton JM, Escalante AA, Neafsey D, Volkman SK. Comparative evolutionary genomics of human malaria parasites. *Trends Parasitol* 2008;**24**:545–50.
42. Pain A, Böhme U, Berry AE, Mungall K, Finn RD, Jackson AP, et al. The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature* 2008;**455**:799–803.
43. Ansari HR, Templeton TJ, Subudhi AK, Ramaprasad A, Tang J, Lu F, et al. Genome-scale comparison of expanded gene families in *Plasmodium ovale wallikeri* and *Plasmodium ovale curtisi* with *Plasmodium malariae* and with other *Plasmodium* species. *Int J Parasitol* 2016 [in press].
44. Robinson T, Campino SG, Auburn S, Assefa SA, Polley SD, Manske M, et al. Drug-resistant genotypes and multi-clonality in *Plasmodium falciparum* analysed by direct genome sequencing from peripheral blood of malaria patients. *PLoS One* 2011;**6**:e23204.
45. Weatherall DJ, Miller LH, Baruch DI, Marsh K, Doumbo OK, Casals-Pascual C, et al. Malaria and the red cell. *Hematol Am Soc Hematol Educ Program* 2002;**2002**:35–57.
46. Winter G, Kawai S, Haeggström M, Kaneko O, von Euler A, Kawazu S, et al. SURFIN is a polymorphic antigen expressed on *Plasmodium falciparum* merozoites and infected erythrocytes. *J Exp Med* 2005;**201**:1853–63.
47. Ochola LI, Tetteh KK, Stewart LB, Riitho V, Marsh K, Conway DJ. Allele frequency-based and polymorphism-versus-divergence indices of balancing selection in a new filtered set of polymorphic genes in *Plasmodium falciparum*. *Mol Biol Evol* 2010;**27**: 2344–51.
48. Weedall GD, Preston BM, Thomas AW, Sutherland CJ, Conway DJ. Differential evidence of natural selection on two leading sporozoite stage malaria vaccine candidate antigens. *Int J Parasitol* 2007;**37**:77–85.
49. Pinheiro MM, Ahmed MA, Millar SB, Sanderson T, Otto TD, Lu WC, et al. *Plasmodium knowlesi* genome sequences from clinical isolates reveal extensive genomic dimorphism. *PLoS One* 2015;**10**:e0121303.
50. Palumbi SR, Metz EC. Strong reproductive isolation between closely related tropical sea urchins (genus *Echinometra*). *Mol Biol Evol* 1991;**8**:227–39.
51. Amato A, Kooistra WH, Ghiron JH, Mann DG, Proschold T, Montresor M. Reproductive isolation among sympatric cryptic species in marine diatoms. *Protist* 2007;**158**:193–7.
52. Perkins SL, Schall JJ. A molecular phylogeny of malarial parasites recovered from cytochrome b gene sequences. *J Parasitol* 2002;**88**:972–8.
53. Mu J, Joy DA, Duan J, Huang Y, Carlton J, Walker J, et al. Host switch leads to emergence of *Plasmodium vivax* malaria in humans. *Mol Biol Evol* 2005;**22**:1686–93.

Downloaded From Vet-eBooks.com

54. Ricklefs RE, Fallon SM. Diversification and host switching in avian malaria parasites. *Proc Biol Sci* 2002;**269**:885–92.
55. Iezhova TA, Valkiunas G, Bairlein F. Vertebrate host specificity of two avian malaria parasites of the subgenus *Novyella*: *Plasmodium nucleophilum* and *Plasmodium vaughani*. *J Parasitol* 2005;**91**:472–4.
56. Duval L, Robert V, Csorba G, Hassanin A, Randrianarivelojosia M, Walston J, et al. Multiple host-switching of Haemosporidia parasites in bats. *Malar J* 2007;**6**:157.
57. Brunet M, Guy F, Pilbeam D, Mackaye HT, Likius A, Ahounta D. A new hominid from the Upper Miocene of Chad Central Africa. *Nature* 2002;**418**:145–51.
58. Hacia JG. Genome of the apes. *Trends Genet* 2001;**17**:637–45.
59. Martin MJ, Rayner JC, Gagneux P, Barnwell JW, Varki A. Evolution of human-chimpanzee differences in malaria susceptibility: relationship to human genetic loss of *N*-glycolylneuraminic acid. *Proc Natl Acad Sci USA* 2005;**102**:12819–24.
60. Lal AA, Goldman IF. Circumsporozoite protein gene from *Plasmodium reichenowi* a chimpanzee malaria parasite evolutionarily related to the human malaria parasite *Plasmodium falciparum*. *J Biol Chem* 1991;**266**:6686–9.
61. Miller LH. Hypothesis on the mechanism of erythrocyte invasion by malaria merozoites. *Bull WHO* 1977;**55**:157–62.
62. Liu W, Li Y, Learn GH, Rudicell RS, Robertson JD, Keele BF, et al. Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. *Nature* 2010;**467**:420–5.
63. Sundararaman SA, Plenderleith LJ, Liu W, Loy DE, Learn GH, Li Y, et al. Genomes of cryptic chimpanzee *Plasmodium* species reveal key evolutionary events leading to human malaria. *Nat Commun* 2016;**7**:11078.
64. Siddiqui WA, Schnell JV, Geiman QM. A model in vitro system to test the susceptibility of human malarial parasites to antimalarial drugs. *Am J Trop Med Hyg* 1972;**21**:393–9.
65. Geiman QM, Meagher MJ. Susceptibility of a New World monkey to *Plasmodium falciparum* from man. *Nature* 1967;**215**:437–9.
66. Hayakawa T, Arisue N, Udono T, Hirai H, Sattabongkot J, Toyama T, et al. Identification of *Plasmodium malariae*, a human malaria parasite, in imported chimpanzees. *PLoS One* 2006;**4**:e7412.
67. Liu W, Li Y, Shaw KS, Learn GH, Plenderleith LJ, Malenke JA, et al. African origin of the malaria parasite *Plasmodium vivax*. *Nat Commun* 2014;**5**:3346.
68. Van Dijk MR, Janse CJ, Thompson J, Waters AP, Braks JA, Dodemont HJ, et al. A central role for P48/45 in malaria parasite male gamete fertility. *Cell* 2001;**104**:153–64.
69. Conway DJ, Machado RL, Singh B, Dessert P, Mikes ZS, Povoa MM, et al. Extreme geographical fixation of variation in the *Plasmodium falciparum* gamete surface protein gene Pfs48/45 compared with microsatellite loci. *Mol Biochem Parasitol* 2001;**115**: 145–56.
70. Anthony TG, Polley SD, Vogler AP, Conway DJ. Evidence of non-neutral polymorphism in *Plasmodium falciparum* gamete surface protein genes Pfs47 and Pfs48/45. *Mol Biochem Parasitol* 2007;**156**:117–23.
71. Oguike MC, Betson M, Burke M, Nolder D, Stothard R, Kleinschmidt I, et al. *Plasmodium ovale curtisi* and *Plasmodium ovale wallikeri* circulate simultaneously in African communities. *Int J Parasitol* 2011;**41**:677–83.
72. Oguike MC, Sutherland CJ. Dimorphism in genes encoding sexual-stage proteins of *Plasmodium ovale curtisi* and *Plasmodium ovale wallikeri*. *Int J Parasitol* 2015;**45**: 449–54.
73. Fuehrer HP, Habler VE, Fally MA, Harl J, Starzengruber P, Swoboda P, et al. *Plasmodium ovale* in Bangladesh: genetic diversity and the first known evidence of the sympatric

Downloaded From Vet-eBooks.com

distribution of *Plasmodium ovale curtisi* and *Plasmodium ovale wallikeri* in southern Asia. *Int J Parasitol* 2012;**42**:693−9.
74. Tanomsing N, Imwong M, Sutherland CJ, Dolecek C, Hien TT, Nosten F, et al. Genetic marker suitable for identification and genotyping of *Plasmodium ovale curtisi* and *Plasmodium ovale wallikeri*. *J Clin Microbiol* 2013;**51**:4213−6.
75. Putaporntip C, Hughes AL, Jongwutiwes S. Low level of sequence diversity at merozoite surface protein-1 locus of *Plasmodium ovale curtisi* and *P. ovale wallikeri* from Thai isolates. *PLoS One* 2013;**8**:e58962.
76. Ahmed AM, Pinheiro MM, Divis PC, Siner A, Zainudin R, Wong IT, et al. Disease progression in *Plasmodium knowlesi* malaria is linked to variation in invasion gene family members. *PLoS Negl Trop Dis* 2014;**8**:e3086.
77. Moon RW, Sharaf H, Hastings CH, Ho YS, Nair MB, Rchiad Z, et al. Normocyte binding protein, NBPXa, is required for human erythrocyte invasion by the zoonotic malaria parasite *Plasmodium knowlesi*. *Proc Natl Acad Sci USA* 2016;**113**:7231−6.
78. Hall N, Karras M, Raine JD, Carlton JM, Kooij TW, Berriman M, et al. A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses. *Science* 2005;**307**:82−6.
79. Krief S, Escalante AA, Pacheco MA, Mugisha L, André C, Halbwax M, et al. On the diversity of malaria parasites in African apes and the origin of *Plasmodium falciparum* from Bonobos. *PLoS Pathog* 2010;**6**:e1000765.
80. Délicat-Loembet L, Rougeron V, Ollomo B, Arnathau C, Roche B, Elguero E, et al. No evidence for ape *Plasmodium* infections in humans in Gabon. *PLoS One* 2015;**10**: e0126933.
81. Makanga B, Yangari P, Rahola N, Rougeron V, Elguero E, Boundenga L, et al. Ape malaria transmission and potential for ape-to-human transfers in Africa. *Proc Natl Acad Sci USA* 2016;**113**. pii:201603008.
82. Zheng H, Zhu HM, Ning BF, Li XY. Molecular identification of naturally acquired *Plasmodium knowlesi* infection in a human case. *Zhongguo Ji Sheng Chong Xue Yu Ji Sheng Chong Bing Za Zhi* 2006;**24**:273−6 [Chinese].
83. Lee KS, Cox-Singh J, Brooke G, Matusop A, Singh B. *Plasmodium knowlesi* from archival blood films: further evidence that human infections are widely distributed and not newly emergent in Malaysian Borneo. *Int J Parasitol* 2009;**39**:1125−8.
84. Escalante AA, Smith DL, Kim Y. The dynamics of mutations associated with anti-malarial drug resistance in *Plasmodium falciparum*. *Trends Parasitol* 2009;**25**:557−63.
85. Mehlotra RK, Mattera G, Bockarie MJ, Maguire JD, Baird JK, Sharma YD, et al. Discordant patterns of genetic variation at two chloroquine resistance loci in worldwide populations of the malaria parasite *Plasmodium falciparum*. *Antimicrob Agents Chemother* 2008;**52**:2212−22.
86. Beshir K, Sutherland CJ, Merinopoulos I, Durrani N, Leslie T, Rowland M, et al. Amodiaquine resistance in *Plasmodium falciparum* malaria is associated with the *pfcrt* 72-76 SVMNT allele in Afghanistan. *Antimicrob Agents Chemother* 2010;**54**:3714−6.
87. Ord R, Alexander N, Dunyo S, Hallett RL, Jawara M, Targett GAT, et al. Seasonal carriage of pfcrt and pfmdr1 alleles in Gambian *Plasmodium falciparum* imply reduced fitness of chloroquine-resistant parasites. *J Infect Dis* 2007;**196**:1613−9.
88. Laufer MK, Thesing PC, Eddington ND, Masonga R, Dzinjalamala FK, Takala SL, et al. Return of chloroquine antimalarial efficacy in Malawi. *N Engl J Med* 2006;**355**:1959−66.
89. Duraisingh MT, Jones P, Sambou I, von Seidlein L, Pinder M, Warhurst DC. The tyrosine-86 allele of the pfmdr1 gene of *Plasmodium falciparum* is associated with increased

Downloaded From Vet-eBooks.com

sensitivity to the anti-malarials mefloquine and artemisinin. *Mol Biochem Parasitol* 2000; **108**:13−23.
90. Humphreys GA, Merinopoulos I, Ahmed J, Whitty CJM, Mutabingwa TK, Sutherland CJ, et al. Amodiaquine and artemether-lumefantrine select distinct alleles of the *Plasmodium falciparum pfmdr1* gene in Tanzanian children treated for uncomplicated malaria *Antimicr. Agents Chemother* 2007;**51**:991−7.
91. Suwanarusk R, Russell B, Chavchich M, Chalfein F, Kenangalem E, Kosaisavee V, et al. Chloroquine resistant *Plasmodium vivax*: in vitro characterisation and association with molecular polymorphisms. *PLoS One* 2007;**2**:e1089.
92. Hunt P, Afonso A, Creasey A, Culleton R, Sidhu AB, Logan J, et al. Gene encoding a deubiquitinating enzyme is mutated in artesunate- and chloroquine-resistant rodent malaria parasites. *Mol Microbiol* 2007;**65**:27−40.
93. Beshir KB, Sutherland CJ, Sawa P, Drakeley CJ, Okell L, Mweresa CK, et al. Residual *Plasmodium falciparum* parasitemia in Kenyan children after artemisinin-combination therapy is associated with increased transmission to mosquitoes and parasite recurrence. *J Infect Dis* 2014;**208**:2017−24.
94. Henriques G, Hallett RL, Beshir KB, Gadalla NB, Johnson RE, Burrow R, et al. Directional selection at the *pfmdr1*, *pfcrt*, *pfubp1*, and *pfap2mu* loci of *Plasmodium falciparum* in Kenyan children treated with ACT. *J Infect Dis* 2014;**210**:2001−8.
95. Spring MD, Lin JT, Manning JE, Vanachayangkul P, Somethy S, Bun R, et al. Dihydroartemisinin-piperaquine failure associated with a triple mutant including kelch13 C580Y in Cambodia: an observational cohort study. *Lancet Infect Dis* 2015;**15**:683−91.
96. Conrad MD, LeClair N, Arinaitwe E, Wanzira H, Kakuru A, Bigira V, et al. Comparative impacts over 5 years of artemisinin-based combination therapies on *Plasmodium falciparum* polymorphisms that modulate drug sensitivity in Ugandan children. *J Infect Dis* 2014;**210**:344−53.
97. Noedl H, Se Y, Schaecher K, Smith BL, Socheat D, Fukuda MM. Artemisinin resistance in Cambodia 1 (ARC1) Study Consortium. Evidence of artemisinin-resistant malaria in western Cambodia. *N Engl J Med* 2008;**359**:2619−20.
98. Dondorp AM, Nosten F, Yi P, Das D, Phyo AP, Tarning J, et al. Artemisinin resistance in *Plasmodium falciparum* malaria. *N Engl J Med* 2013;**361**:455−67.
99. Ye R, Hu D, Zhang Y, Huang Y, Sun X, Wang J, et al. Distinctive origin of artemisinin-resistant *Plasmodium falciparum* on the China-Myanmar border. *Sci Rep* 2016;**6**:20100.
100. Muwanguzi J, Henriques G, Sawa P, Bousema T, Sutherland CJ, Beshir KB. Lack of K13 mutations in *Plasmodium falciparum* persisting after artemisinin combination therapy treatment of Kenyan children. *Malar J* 2016;**15**:36.
101. Takala SL, Coulibaly D, Thera MA, Dicko A, Smith DL, Guindo AB, et al. Dynamics of polymorphism in a malaria vaccine antigen at a vaccine-testing site in Mali. *PLoS Med* 2007;**4**:e93.
102. Sutherland C. A challenge for the development of malaria vaccines: polymorphic target antigens. *PLoS Med* 2007;**4**:e116.
103. Neafsey DE, Juraska M, Bedford T, Benkeser D, Valim C, Griggs A, et al. Genetic diversity and protective efficacy of the RTS,S/AS01 malaria vaccine. *N Engl J Med* 2015; **373**:2025−37.
104. Richie TL, Billingsley PF, Sim BK, James ER, Chakravarty S, Epstein JE, et al. Progress with *Plasmodium falciparum* sporozoite (PfSPZ)-based malaria vaccines. *Vaccine* 2015; **33**:7452−61.

Downloaded From Vet-eBooks.com

This page intentionally left blank

Downloaded From Vet-eBooks.com

Integrated Genetic Epidemiology of Chagas Disease

22

M. Tibayrenc[1], *M.A. Shaw*[2]
[1]Maladies Infectieuses et Vecteurs Ecologie, Génétique, Evolution et Contrôle MIVEGEC (IRD 224-CNRS 5290-UM1-UM2), IRD, Montpellier, France; [2]University of Leeds, St James's University Hospital, Leeds, United Kingdom

1. What Is Integrated Genetic Epidemiology?

As authoritatively illustrated by this book, the impressive progress of molecular megatechnologies (high-throughput sequencing, microarrays, postgenomics) and the concomitant development of bioinformatics have considerably improved our knowledge on infectious diseases. However, there is a strong tendency toward compartmentalization in the research effort: scientists working on human (and other hosts) genetic susceptibility to infectious diseases are generally not aware of research on the role played by pathogens and vectors in the case of vector-borne diseases. This results in each community of scientists tending to overemphasize the role of its study material. This compartmentalization is all the more distressing since coevolution between hosts, pathogens, and vectors should be considered a unique biological phenomenon. The term "integrated genetic epidemiology"[1] has been coined to designate the approach consisting in simultaneously analyzing the impact of the host's, the pathogen's, and the vector's genetic diversity on the transmission and severity of infectious diseases as well as the coevolution processes between the three. The present chapter aims to show that Chagas disease (CD) is an excellent model to develop this approach. It briefly summarizes what is presently known about (1) human genetic susceptibility to CD, (2) the vectors' species and population diversity, and (3) the parasite's genetics and evolution. Then it demonstrates how these three components could be merged in a unique approach.

2. Chagas Disease: A Major Health Problem in Latin America and Other Countries

CD remains by far the most serious health problem in Latin America. Control of the disease has been improved, but several million people remain at risk or are stricken by the disease.

From a clinical point of view, CD is a very serious illness. After infection by the parasite (see Section 3), patients develop an acute phase, which actually corresponds to parasitic septicemia. Mortality at this stage is about 5%. After a few weeks, patients

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00022-6
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

who survive enter the indeterminate phase, with no symptoms. About 70% of patients will never exhibit any symptoms again. However, 30% of them will develop symptomatic CD. The most worrisome symptom is Chagasic cardiopathy, which leads to a severe cardiac insufficiency. Other clinical forms involve the digestive system (megacolon, megaesophagus) and cause severe functional abnormalities.

The health problem of CD is worsened by its being a "neglected disease" according to official classifications. As for the infectious diseases predominant in the South, malaria, AIDS, and tuberculosis receive special attention from WHO and other international health authorities, while other diseases tend to be underprioritized. However, the dispersion of CD from Latin America to nonendemic countries by population movements has started to create new epidemiological, economic, social, and political challenges as *Trypanosoma* cruzi has spread throughout the world[2]. In the domain of scientific research, it is notable that the scientific community involved in CD research, although very productive, is tiny, but hopefully will expand.

3. The Chagas Disease Cycle

The CD cycle is only briefly summarized here, since this chapter is not intended to be an exhaustive review of what CD is, but rather attempts to explain why this disease is a valuable model for integrated genetic epidemiology.

The causative agent of CD, a parasitic protozoan of the family Kinetoplastidae, which also includes *Trypanosoma brucei*, the agent of sleeping sickness (African trypanosomiasis) and the *Leishmania*, agents of the various forms of leishmaniosis.

T. cruzi is transmitted by "true" bugs, heteropterous insects of the family Reduviidae, subfamily Triatominae. This subfamily has specialized in obligatory blood feeding. It is worth noting that Chagas vectors include many different species and three principal genera, namely *Triatoma*, *Rhodnius*, and *Panstrongylus*. Vectors are infected by ingesting blood that contains the parasite. They transmit the parasite, not through their biting, but by their feces, which contains the infecting forms. Most vector species present the particularity of depositing feces while they feed on their host. The parasite enters the host by excoriations, through the mucosa, even through intact skin.

Hosts comprise virtually all mammalian species, either domestic or sylvatic, including of course humans.

4. Host Genetic Susceptibility to Chagas Disease

Host, vector, and pathogen genetic diversity will differ in the magnitude of their influence on the outcome of vector-borne infectious diseases. There are indications of a human genetic component to CD from mouse models, and evidence from testing specific human genes. The contribution of human genetics to phenotypic variability is measured as heritability, although for CD there is a paucity of such "family studies."

There are a number of phenotypes that have been used in genetic studies of CD. Investigations have concerned genetic control of infection per se, and also genetic

Downloaded From Vet-eBooks.com

control of chronic disease phenotypes. An individual defined as resistant must necessarily have been exposed, and many studies employ anti-*T. cruzi* antibody levels to classify individuals as seropositive or seronegative. Those who are seropositive will have had the acute phase and entered into the chronic phase of disease. Some studies, less satisfactorily, use "healthy controls" rather than individuals tested as seronegative, although both groups are relying on individuals being exposed. Of those individuals for whom disease progresses from a quiescent phase and enters into the chronic phase, the most common phenotype studied is that of cardiomyopathy (chronic Chagas cardiomyopathy, CCC). Other phenotypes used for the chronic forms of disease include digestive forms and mixed cardiomyopathic and digestive forms. A phenotype that has received little attention to date, which we might speculate has a genetic component to susceptibility, is congenital CD. However, the vast majority of the work on susceptibility to CD and disease progression, and the work cited here, have been carried out using qualitative, usually dichotomous, traits such as seropositivity versus seronegativity, or presence versus absence of CCC.

There were early estimates of a heritable component to immunoglobulin levels in a CD infected Brazilian population,[3] and an effect of sibling history on Chagas-associated cardiopathy.[4] For CD, while there are uninfected individuals in areas with a high prevalence of disease, this is not necessarily due to genetic resistance and environmental risk factors require consideration. In a study of seropositivity for *T. cruzi* in 716 Brazilian adults, 525 of whom were assigned to 146 pedigrees, an estimate of the heritability of infection of 56% was obtained, with a further 23% of the variation due to shared environment/common household.[5] Another Brazilian study reported on 41 families with 526 individuals found evidence of familial clustering of seropositivity to *T. cruzi*, with 15 families showing seropositivity in >50% of individuals.[6] A sporadic model of seropositivity was clearly rejected, although the causes of familial aggregation could not be established.

Some of the best supported estimates of heritability for infectious diseases, other than CD, do not concern genetic control of infection per se but particular outcomes of infection, sometimes years later. Whilst most CD work has focussed on qualitative traits, Williams-Blangero et al. have analyzed quantitative traits to determine heritabilities for antibody response to *T. cruzi* and traits obtained from ECGs, in an ongoing longitudinal study of >1300 individuals.[7] The heritability for seropositivity to *T. cruzi* rose from 56%, obtained using a dichotomous trait, to 64%. Electrocardiography (ECG)-related traits were also highly heritable, though surprisingly less heritable than seropositivity. There was no good evidence that genes controlling antibody response to *T. cruzi* also controlled cardiovascular traits.[7]

In the context of heritability of susceptibility to infectious disease per se, an estimate of 64% is very high. However, sometimes such high heritabilities have been reported for immunoglobulin/antibody levels, and a high estimate gives some cause for optimism that a hunt for genes might be worthwhile.

4.1 Candidate Gene Approach

This is not an exhaustive review of all published association studies, which were reviewed in early 2010s.[8–10] Most publications on CD are candidate gene studies

Downloaded From Vet-eBooks.com

with often a single-nucleotide polymorphism (SNP) tested per gene. Where the SNP is highly informative (highly heterozygous), and there is functional evidence that the SNP itself is functionally relevant, good information has been obtained. Nevertheless, some studies use a single SNP with poor information content, and without evidence of functionality or linkage disequilibrium with a functional variant (LD), cannot be taken to be conclusive. Many studies have had small sample sizes and while there are a number of reported associations in the literature, not all are statistically significant. One additional consideration for CD is the heterogeneity of the populations tested. Candidate genes have included genes involved in innate and adaptive immunity, coding for signaling molecules and receptors, those involved in signal transduction and cell migration, and genes determining immunomodulatory molecules. This topic has received attention for over 30 years, but the number of published studies is still relatively small when compared to other chronic conditions such as leprosy, and there are insufficient studies for meta-analysis in most instances.

4.2 Classical HLA Associations

Genes in the classical MHC regions typically code for antigen-presenting molecules, highly relevant to development of an appropriate immune response. Class I and class II molecules present *T. cruzi* antigens and heterozygosity at these loci is likely to be beneficial. The principle loci receiving attention are *HLA-A*, *HLA-B*, and *HLA-C* for class I, and *HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1*, and *HLA-DPB1* for class II. Most genotyping has been carried out using sequence-specific oligonucleotide (SSO) and sequence-specific primer (SSP) technologies, and results are published for a number of populations.

Associations are reported for class I alleles and susceptibility/protection versus infection per se,[11–13] for class I alleles and susceptibility/protection versus CCC,[12,14–17] for class II alleles and susceptibility/protection versus infection per se[10,12,16–21] and for class II alleles and susceptibility/protection versus CCC.[12,13,15,18,19,22] There is a minority of studies including digestive and mixed pathologies.[13,15]

While there are more publications for the MHC than for other loci, there are still relatively few, and they are not sufficiently comprehensive for our understanding of reported associations across diverse populations. Many of the associations have not been repeated, and small sample numbers used in some instances are particularly limiting for these highly polymorphic loci. Due the number of genes of interest within the MHC, one or many of which may be relevant to susceptibility, and LD, interpretation of findings may be difficult. Associations need to be considered in depth to pinpoint the primary association, and in this context, some studies have considered combinations of alleles or haplotypes.[14,15,21] More comprehensive studies are needed.

4.3 Further MHC Associations

The only two loci within the MHC class III region with more than one report in the literature are tumor necrosis factor (*TNF*) and lymphotoxin-alpha (*LTA*). *TNF*, coding for the

Downloaded From Vet-eBooks.com

proinflammatory cytokine TNF-a, is perhaps the most widely studied cytokine gene for multifactorial diseases with an immune etiology. *TNF* and *LTA* on chromosome 6 have an extensive infectious disease literature. Most studies on *TNF* and CD concern susceptibility to CCC and test SNPs in the promoter region at positions −1031, −308, and −238.[23−30] The only use of survival analysis for CD is on a small Brazilian sample set and *TNF*.[23] All bar one study investigate TNF-308 despite the poor information content of this marker which, together with relatively small sample sizes used, may contribute to the conflicting *TNF* polymorphism associations reported across populations. If there are associations, then the relative risk of disease afforded by carriage of particular *TNF* alleles is low. A small meta-analysis has been described.[8]

LTA is less investigated than *TNF*, despite the fact that the commonly used *LTA+252* polymorphism is more informative than the *TNF-308* SNP and the two are in LD.[30−32] Because of this LD, studies of both *TNF* and *LTA* are often accompanied by the measurement of TNF-a production. The two main reports provide some indication of association for *LTA+252*G* with CCC.[31,32]

There is also LD between class III loci and genes in the class I and class II regions. The 21-hydroxylase allele *CYP21A2*15* was found to be in strong LD with an HLA resistance haplotype *B*1402-DRB1*0102* (31). Since the same reporting group had previously reported *HLA-B*14-DRB1*01* to be associated with resistance to chronic CD phenotypes, a new primary contribution to resistance of *CYP21A2*15* was ruled out.[15,33]

4.4 Cytokine and Cytokine Receptor Genes

Investigations of cytokine genes often rely on prior knowledge obtained from human immunological studies. In simple terms, a Th1 response predominates in the acute phase of disease, whereas in the chronic phase, both Th1 and Th2 responses are evident, with Th2 response associated with a better outcome. With the exception of one sequence-based study on *IL4*,[34] investigations have used SNPs, and in some instances, only a single SNP is employed. Comparisons have been drawn between seronegatives, asymptomatics, and CCC patients, but there are no studies of digestive or mixed cases.

There are two reports on the IL1 gene cluster, which includes *IL1A*, *IL1B*, and *IL1RN*,[35,36] the larger of which found an increase in the G allele and GG genotype for *IL1B*+5810, and *IL1B* haplotypes with increased risk of CCC.[35] The paucity of studies for this proinflammatory cytokine is unexpected, since for other infectious diseases reports often mirror the abundance of those for *TNF* in the MHC class III region.

Some of the loci tested relate to Th1 responsiveness and IFN-g production such as *IL18*.[37,38] Surprisingly, there is only a single study on *IFNG* with one SNP tested in a Colombian population.[39] The A allele at position 874, associated with reduced IFN-g production, was at a higher frequency in patients compared to seronegatives, although there was no difference in frequency between asymptomatics and cardiomyopathy patients.[39] Genetic studies on cytokines have benefitted from prior work on mouse models where knockout mice are available and have been infected with *T. cruzi*. A good example is a large Colombian study on *IL17A* testing five SNPs and finding associations with disease per se and CCC.[40]

Downloaded From Vet-eBooks.com

However, any gene, whether affiliated to Th1 or Th2 responsiveness, could be contributory to susceptibility since alleles at any one locus may determine high or low levels of a particular cytokine with consequences for the immune response. Since for cytokines per se it is likely that levels of production will influence disease outcome, variants tested often lie within promoter regions, perhaps determining levels of cytokine production themselves, rather than in coding regions lying in LD with causative variants. Accompanying functional studies are often reported in this context, although genetic background is sufficiently diverse that it is hard to draw firm conclusions, and individuals infected with *T. cruzi* may also be mounting immune responses to other infecting pathogens.

Genes coding for cytokines, other than those in the IFN-g pathway have been examined. Although TGFb1 is typically described as a "down-regulatory" cytokine, its role in the development and progression of CD is complex. A thorough study of *TGFB1* using five functionally relevant SNPs found association of a single SNP with susceptibility to disease per se.[41]

Perhaps surprisingly, there is a single reported study testing for the influence of a cytokine receptor gene *IL4RA*[42] with one of four SNPs showing a weak association in a comparison of cardiomyopathic versus asymptomatic patients. This gene has been widely studied in the context of asthma and allergy. The two loci coding for chains of the IFN-g receptor might be well worth study. Very rare variants in these loci, and loci such as *IL12B*, which has been shown to be associated with CCC, cause very severe phenotypes.[43] It has been speculated that rare variants cause rare severe phenotypes as single-gene disorders, whereas common variants in the same loci may control susceptibility/resistance to more common disorders with a genetic component such as CD.

4.5 Chemokine and Chemokine Receptor Genes

The only chemokine/chemokine receptor gene as the subject of more than two reports is *CCR5*.[44–50] *CCR5* is a good example of a candidate gene selected for potential relevance to a particular pathology, CCC, and all reports compare frequencies in seronegatives or asymptomatic patients with CCC. Most commonly tested are the D32 variant, widely studied in the context of HIV1 susceptibility, and the 59029 promoter polymorphism. While the *CCR5D32* can prove uninformative, there is some evidence for association with the promoter polymorphism.[44–46] Suggestion of an effect of variation in *CCL2* with CCC also merits further investigation.[49,51]

4.6 Associations and Other Genes

Especially, in the context of integrated genetic epidemiology, some interesting loci for investigation are those coding for proteins that directly interact with *T. cruzi*. *T. cruzi* is known to activate the TLR system, especially TLR2 and TLR4. *TLR2* and *TLR4*, when represented by a single SNP, failed to show association with susceptibility to disease per se or to CCC, *TLR2* in three studies and *TLR4* in two of the same three studies, although the information content was low.[52–54] However, there is evidence that

Downloaded From Vet-eBooks.com

MAL/TIRAP coding for an adaptor protein involved in the downstream signaling from TLR2 and TLR4 may be associated. Heterozygosity for the S180L variant was associated with a low risk of developing CCC,[53] and homozygosity for 1 of 5 tagSNPs in the 3′UTR in LD with the S180L variant supported this finding.[49] Nevertheless support was not maintained by a 2012 small study.[52] The full biological basis of these observations has yet to be evaluated.

Mannose-binding lectin (MBL) is a soluble pattern recognition molecule that binds to sugar residues on *T. cruzi*, and other pathogens, activating complement. The genetics of MBL has a long history and protein levels are linked to SNPs in *MBL2*. There is one small Chilean study (50) and one larger Brazilian study[55] of the consequences of variation in *MBL2* after infection with *T. cruzi*, the former using SNPs reporting association with disease per se (50), and the latter a sequence approach reporting protection of genetically controlled MBL deficiency against the development and progression of CD.[55] Although both studies report association, *MBL2*C*, a low producing allele, protects against CD and is absent in CCC patients in the larger study,[55] whereas *MBL2*B*, a low producing allele, has a higher frequency in CD patients compared to healthy controls in the smaller study.[52] MBL activates complement through the MBL-associated serine protease MASP2 and the same Brazilian group has also studied six SNPs in *MASP2*[56] Haplotypes were related to MASP-2 levels and MBL/MASP-2/C4 complexes in the context of CCC. Despite positive findings of genotypes with low MASP-2 levels associated with CCC, it remains difficult to relate SNPs to expression levels on diverse haplotypic backgrounds across populations. Similar problems have occurred in other genes, most notably *TNF*.

Also related to the complement system is L-ficolin, encoded by *FCN2*. L-Ficolin binds to acetylated sugars on *T. cruzi*, enhancing both phagocytosis and lysis through complement. A study on Brazilian patients with a range of disease phenotypes looked at four variants within *FCN2*[57] Associations were tested and interpreted together with L-ficolin levels in patients. Whether there is prognostic value in complex associations remains to be established. Nevertheless further investigation is warranted.

Two immunomodulatory molecules that have each been studied in two populations are CTLA-4, coded for by *CTLA4*[58,59] and haptoglobin, coded for by *HP*[60,61] but both with mixed results. As for the cytokine genes, among these other genes examined, protein expression levels are often related to disease outcome, although causality is not established. Brazilian studies have considered patients with digestive and mixed phenotypes, as well as those with CCC.[55–58]

4.7 A Genome-Wide Approach

Across all multifactorial disorders, the period 2000s have seen a rapid rise in the number of publications of genome-wide association studies (GWAS) from use of SNP chips, with association relying on LD between genotyped SNPs and causal variants. Studies are mainly case–control design. Replication is key and a good proportion of GWAS should incorporate a second dataset to test for reproducibility. There are a number of sites cataloging GWAS results to enable

Downloaded From Vet-eBooks.com

scientists to see genes highlighted in multiple studies (GWAS Central at www.gwascentral.org and https://www.ebi.ac.uk/gwas/). Diseases with an immune component often "share" genes with other conditions, and results from GWAS may stimulate more targeted candidate gene approaches. Issues, such as population heterogeneity, are common to both GWAS and candidate gene studies. The GWAS approach is described further by Bush et al.[62] and the first five years reflected on by Visscher et al.[63]

There is a single reported GWAS for CD, specifically for cardiomyopathy in *T. cruzi* seropositive subjects.[64] Of the 600 Brazilian samples tested, 221 were classified as CCC, 311 had no cardiomyopathy, and 68 were inconclusive. Genotyping used an Affymetrix array with more than 800,000 SNPs, with a further 5 million SNPs imputed. Seven phenotypes were analyzed including anti-*T. cruzi* antibody levels, cardiomyopathy and parameters from ECG. For cardiomyopathy, the final analysis was on 207 CCC and 306 non-CCC samples with population admixture quantified for each individual. Two SNPs were highlighted in *SLCO1B1,* coding for a solute carrier that plays a role in drug metabolism, with respect to cardiomyopathy. A total of 46 SNPs in novel genes were described as associated with the seven traits, but none of the SNPs reached accepted genome-wide significance levels and this is almost certainly due to the low power of this dataset. More typical numbers for a GWAS would be 2000 cases and 2000 controls, with additional numbers for a replicate dataset. Any further work will rely on a collaborative approach.

4.8 The Future

Despite nearly 30 years of interest in the genetics of susceptibility to CD per se and chronic disease such as CCC, we are still in the early stages of investigation. Although estimates of heritability are encouraging that the search for genes contributing to disease susceptibility will be possible and prove useful, the numbers of studies indicating a genetic component to phenotypic variance are few.

Candidate gene studies are providing some clues as to potentially contributory loci. Nevertheless, ethnic diversity in regions where CD is endemic, and the problems of collecting suitable numbers of samples, when chronic disease may be slow to develop, are evident. There is a need for more studies with greater numbers of samples and polymorphisms. Changing laboratory technologies should enable investigators to replicate and expand the number of candidate genes considered, improving coverage of these loci and, with this, new questions will be asked.

Undoubtedly larger GWASs would be worthwhile. Since GWASs often detect new loci, but have been known to miss proven susceptibility loci, candidate gene and GWAS approaches can be regarded as complementary. However, there are few strong candidate genes for CD, it appears unlikely that a single gene will account for a large proportion of the heritability and the risk conferred by each contributing locus will be small. Positive associations are likely to provide information on the biology of CD and should contribute to our understanding of the epidemiology.

Downloaded From Vet-eBooks.com

5. Vector Genetic Diversity

CD exhibits a specific epidemiological feature, namely that the parasite can be transmitted by an impressive range of different vectors. They all pertain to the category of "true bugs" (order Hemiptera, suborder Heteroptera). They are all included in the subfamily Triatominae, family Reduviidae. While other Reduviidae are predators, the Triatominae have specialized in obligatory blood feeding, including adults of both sexes and larvae. Within the subfamily Triatominae, three main genera of unequal ecogeographic distribution can transmit CD, namely *Triatoma*, *Rhodnius,* and *Panstrongylus*. Each of these genera includes various species that are able to transmit the disease.

The genetic diversity of the vectors at both the genera and the species levels is therefore considerable (see Chapter 15, this volume, for more information).

At the subspecific level, many studies have explored the diversity of many species, both by population genetic markers (see Chapter 15, this volume) and by computer-assisted morphometric analysis; therefore, the diversity of CD vectors at the subspecific and population level is fairly well known.

However, little is known about the differential vectorial capacity of the various triatomine species and of different populations within species. The null hypothesis that all species and all populations of a given species are equally able to transmit *T. cruzi* and its various genotypes (see Section 6) can be ruled out. It is highly conceivable that refined coevolution phenomena have occurred, meaning that local vectors are better able to transmit local parasite genotypes. This remains to be explored. It is worth noting, however, that a North American vector (*Triatoma protracta*) is fully able to transmit a Latin American strain of *T. cruzi* in experimental conditions.[65]

6. Parasite Genetic Diversity

It is interesting to note that, although the scientific community working on CD is small compared to the numbers working on AIDS, malaria, or tuberculosis, this pathogen has long been among the pioneer species explored by advanced approaches such as molecular typing and population genetics. Therefore, this parasite is probably the pathogen whose evolutionary biology is the best known, together with *Escherichia coli*. It can therefore be suggested as a paradigmatic biological model, as has been done with *E. coli*, *Drosophila melanogaster*, *Mus musculus*, and *Caenorhabditis elegans*.[66]

Pioneering molecular studies on *T. cruzi* explored *isoenzyme* variability as early as the beginning of the 1970s.[67] Although a now out-of-fashion technique, *multilocus enzyme electrophoresis* has clearly discriminated three principal variants or zymodemes within *T. cruzi*.[68] It is interesting to note that this observation remains current, since these three zymodemes continue to be recorded today in *T. cruzi* natural populations, although their denomination and evolutionary status has changed substantially. This permanency of *multilocus genotypes* over space and time is one of the strongest arguments in favor of predominant clonal evolution (PCE; see later).

Downloaded From Vet-eBooks.com

The interpretation of isoenzyme diversity in terms of population genetics and evolutionary biology has made it possible to clarify the evolutionary status of the zymodemes. The model of PCE has been proposed for *T. cruzi*[69] and for other parasitic protozoa.[70] Recent developments have made it possible to extend the PCE model, from eukaryotic pathogens (parasitic protozoa and fungi)[71] to bacteria and viruses.[72] The evidence for PCE in *T. cruzi* was mainly based on the observation of a considerable linkage disequilibrium (nonrandom association of genotypes occurring at different loci). Linkage disequilibrium is the very manifestation of very limited or absent genetic *recombination*. The PCE model stipulates that offspring multilocus genotypes are virtually identical to the parental genotypes and are stable in space and time, whatever the precise cytological mechanism of propagation. The model therefore includes not only mitotic propagation, but also various forms of parthenogenesis, extreme homogamy, and self-fertilization in haploid organisms.[70] Extreme inbreeding is not an alternative model to clonal evolution,[73,74] but rather a particular case of it.[72]

The main relevance of the model concerns molecular epidemiology (tracing multilocus genotypes (strains) with molecular tools for epidemiological follow-up). If PCE inhibits recombination, as stated earlier, the multilocus genotypes are stable in space and time, even at an evolutionary scale, and therefore constitute convenient targets for molecular epidemiology.

Since its inception, the clonal model has stated that it was compatible with occasional bouts of genetic recombination. Recombination has long been suspected in natural populations of *T. cruzi*[75] and has been experimentally evidenced.[76] However, it is clear that such hybridization events interfere only at an evolutionary scale. The stability of *T. cruzi* multilocus genotypes in the long run, with its extreme manifestation of strong parity between phylogenetic trees designed from different genetic markers[77] is incompatible with frequent genetic recombination.

It has been suggested that *T. cruzi* genotypes are distributed into six different clusters[78,79] (Fig. 22.1), which cannot be equated with real *clades* because some of them clearly originate from former hybridization events,[80,81] further stabilized by clonal propagation. The term "discrete typing unit" (DTU)[1] has been coined to designate sets of stocks that are genetically closer to each other than to any other stock and are identifiable by common molecular, genetic, biochemical, or immunological markers called tags. The six *T. cruzi* clusters match this definition. Their validity was confirmed at a meeting of CD experts.[82,83]

In the light of these developments of the PCE model,[72] we have proposed that one of the main consequences of PCE was the generation of "near-clades," or clades which discreteness is somewhat clouded by occasional hybridization/genetic exchange. Since occasional genetic exchange almost always is observed in pathogenic microorganisms, too demanding cladistic criteria are inappropriate to explore their subspecific genetic variability. However, the presence of near-clades can be conveniently evidenced by a flexible phylogenetic approach relaxing cladistic demands. Such an approach is based on a congruence criterion inspired from the principle of genealogical concordance between independent genes proposed for the recognition of biological taxa.[84] According to this congruence criterion, adding more relevant data (e.g., more loci,

Downloaded From Vet-eBooks.com

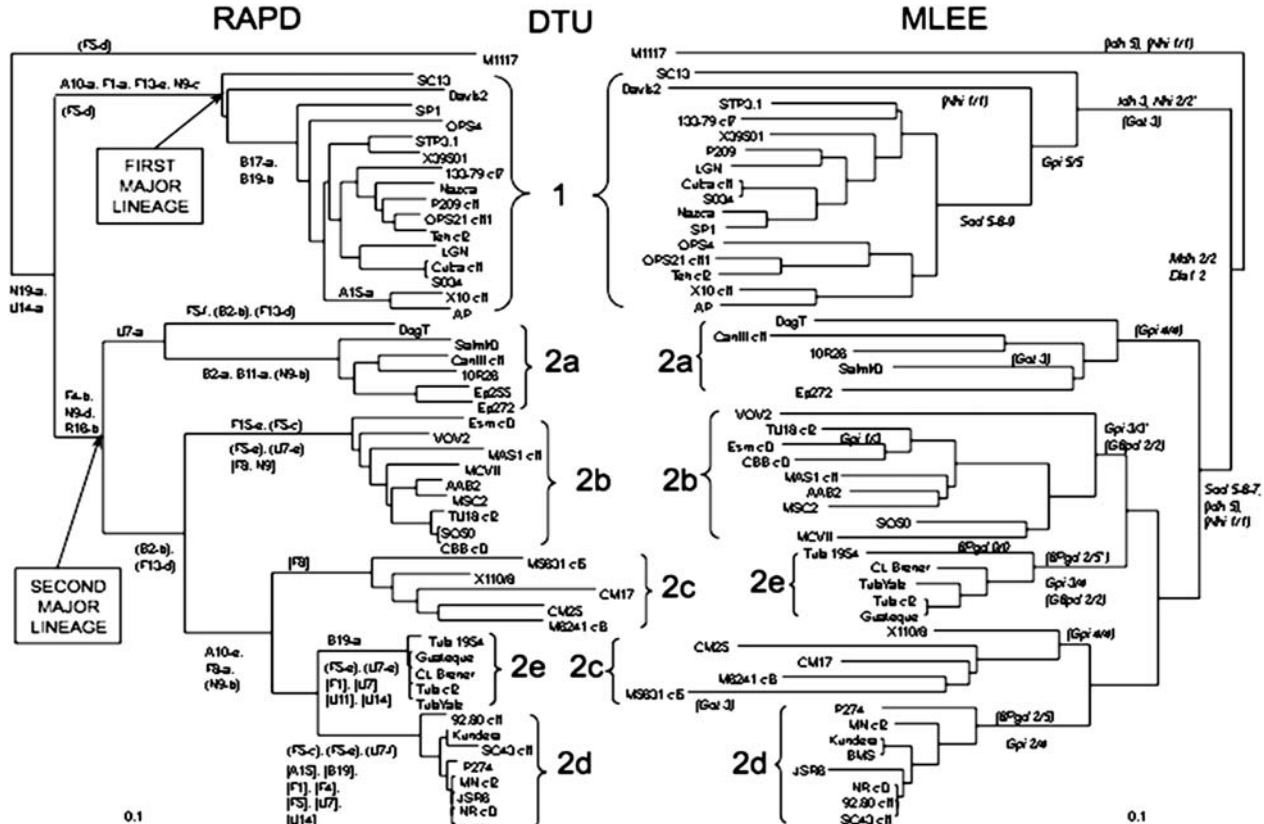


Figure 22.1 Double phylogenetic tree depicting the evolutionary relationships among *Trypanosoma cruzi* genotypes: isoenzymes (left) and RAPD (right). The fair parity between the two trees is a clear evidence of linkage disequilibrium and predominant clonal evolution.
From Brisse S, Barnabé C, Tibayrenc M. Identification of six *Trypanosoma cruzi* phylogenetic lineages by random amplified polymorphic DNA and multilocus enzyme electrophoresis. *Int J Parasitol* 2000;**30**:35−44.

Downloaded From Vet-eBooks.com

or more molecular markers, or data obtained from different phylogenetic approaches), will evidence an increasing phylogenetic signal in the population under study. We have proposed that such a growing phylogenetic signal, which is easy to evidence with appropriate data, is the criterion for defining a "clonality threshold," beyond which the impact of clonal evolution definitely surpasses that of genetic recombination.[85] The clonality threshold concept makes it possible to abandon vague, subjective terms such as "gross" incongruences (between phylogenetic trees),[86] "widespread genetic exchange,"[87] "intense lateral exchange of genetic information,"[88] and others, and to rather rely on a clear—cut parameter that identifies the "clonality border." *T. cruzi* DTUs perfectly fit the definition of near-clades. The near-clade concept, by comparison with the DTU concept, presents the advantage of having a clear evolutionary meaning. It is widely applicable to many pathogenic microorganisms, including not only parasites, but also fungi, bacteria, and viruses.[72]

A seventh *T. cruzi* near-clade, referred to as "Tc-Bat" (because it has been observed only in bats) has been recently evidenced.[89–92] Tc-bat has been recorded in Brazil, Colombia, Ecuador, and Panama years apart, in various bat species, which illustrates the stability in space and time of the near-clades. Lastly, the classification into seven near-clades was challenged[93] on a broad sample of strains, but with a limited set of

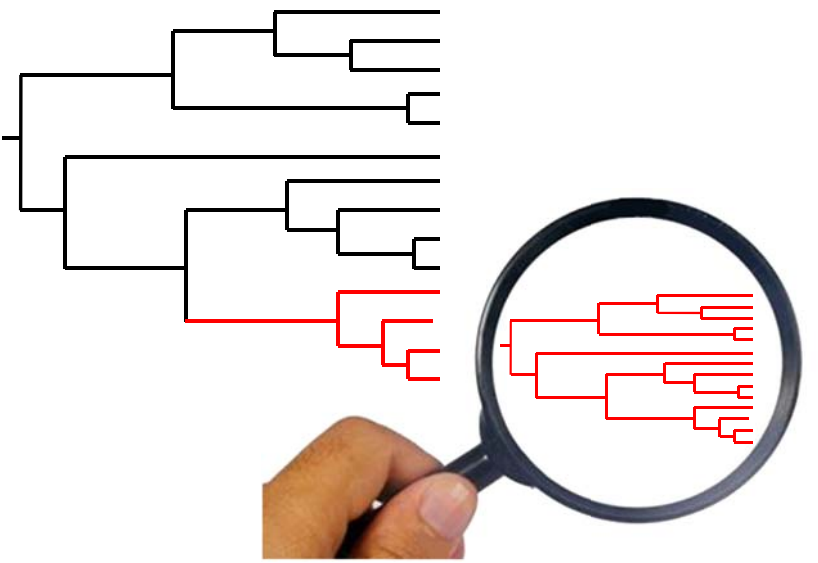


Figure 22.2 "Russian doll" pattern. When population genetic tests are performed with adequate markers (of sufficient resolution) within each of the near-clades that subdivide the species under study (large tree, left part of the figure), they evidence a miniature picture of the whole species, with the two main PCE features, namely linkage disequilibrium and lesser near-clades (small tree, right part of the figure). This is evidence that the near-clades are not cryptic biological species, and that they also undergo predominant clonal evolution.

Downloaded From Vet-eBooks.com

genetic markers (one nuclear gene and two mitochondrial genes). This proposal should obviously be explored with a broader set of genetic markers. However, it clearly shows that even in the extremely well-studied species *T. cruzi*, evidencing near-clades is far from being "self-evident."[88]

Developments of the PCE model in 2013 shown a "Russian doll pattern" in *T. cruzi*, as well as in other microorganisms,[94] that is to say: within the near-clades, PCE also operates, and leads to a within-near clade miniature picture of the population structure of the whole species, with lesser near-clades, linkage disequilibrium and clonal population structure (Fig. 22.2).

From the points of view of molecular epidemiology, experimental evolution, and integrated genetic epidemiology, the population structure of *T. cruzi* summarized earlier can be illustrated by two keywords: stability and discreteness. *T. cruzi* natural clones, and the near-clades into which they are distributed, are genetic entities that are both stable in space and time (up to the evolutionary scale) and strictly separated from each other, with rare occasional bouts of genetic exchange. *T. cruzi* near-clades have been taken as units of analysis to explore the variability of experimental parameters[95] and the differential expression of genes through proteomic analysis.[96,97]

7. Concluding Remarks

The data described herein were not intended to be a comprehensive review of our present knowledge on the genetic diversity of CD hosts, vectors, and parasites. Instead, the goal was to briefly highlight why these data make CD a good model for the integrated genetic epidemiology of infectious diseases, as already proposed long ago.[98]

The keyword here again is discreteness: discreteness of the clinical phenotypes of CD in humans, discreteness of *T. cruzi* clonal genotypes and near-clades, discreteness of the many different species that are hosts (mammals) and vectors (triatomine bugs) of CD. All these discrete entities can be used as units of analysis, keys on the keyboard to be played in many different situations that can be analyzed, both in surveying natural Chagas cycles and in designing experimental evolution protocols.

There are several possible examples.

When natural cycles are considered, possible protocols could be to compare *T. cruzi* genotypes isolated from (1) cardiac versus digestive versus asymptomatic patients; (2) different mammal species; and (3) different triatomine bug species.

Experimental evolution protocols are easy because (1) *T. cruzi* is easy to culture; (2) many triatomine bug species are easy to raise; and (3) several experimental animal models are available, and one can compare, for example, different breeds of mice, whose genetic distinctness results in differing susceptibility to CD.

All this makes the integrated genetic epidemiology of CD an extremely promising field of research that has until now been underexplored. It could constitute a paradigmatic example to develop similar approaches in other infectious models.

Downloaded From Vet-eBooks.com

Glossary

Clade Evolutionary lineage defined by cladistic analysis. A clade is monophyletic (it has only a single ancestor) and is genetically isolated (which means that it evolves independently) from other clades.

Isoenzymes, multilocus enzyme electrophoresis Protein extracts of given biological samples are separated by electrophoresis. The gel is then processed with a biochemical reaction involving the specific substrate of a given enzyme. This enzyme's zone of activity is then specifically stained on the gel. From one sample to another, migration differences can appear for this same enzyme. These different electrophoretic forms of the same enzyme are referred to as isoenzymes or isozymes. These differences reflect sequence differences in the genes coding for the involved enzymes.

Multilocus genotype The combined genotype of a given strain or a given individual established with several genetic loci.

Phenotype All observable properties of a given individual or a given population apart from the genotype. The phenotype is not limited to morphological characteristics and can include, for example, physiological or biochemical parameters. The pathogenicity of a microorganism is a phenotypic property, as are the different clinical forms of a given disease. The phenotype is produced by the interaction between genotype and the environment.

Phylogeny, phylogenetic Evolutionary relationships among taxa, species, organisms, genes, or molecules.

Population genetics Analysis of allele and genotype frequency distribution and modifications under the influence of natural selection, mutation, genetic drift, and gene flow.

References

1. Tibayrenc M. Genetic epidemiology of parasitic protozoa and other infectious agents: the need for an integrated approach. *Int J Parasitol* 1998;**28**:85−104.
2. Rodrigues Coura J, Albajar Viñas P. Chagas disease: a new world challenge. *Nature* 2010; **465**:S6−7.
3. Barbosa CAA, Morton NE, Rao DC, Krieger H. Biological and cultural determinants of immunoglobulin levels in a Brazilian population with Chagas-disease. *Hum Genet* 1981;**59**: 161−3.
4. Zicker F, Smith PG, Netto JCA, Oliveira RM, Zicker EMS. Physical-activity opportunity for reinfection, and sibling history of heart-disease as risk-factors for Chagas cardiopathy. *Am J Trop Med Hyg* 1990;**43**:498−505.
5. Williams-Blangero S, Vandeberg JL, Blangero J, Teixeira ARL. Genetic epidemiology of seropositivity for *Trypanosoma cruzi* infection in rural Goias, Brazil. *Am J Trop Med Hyg* 1997;**57**:538−43.
6. Silva-Grecco RL, Balarin MAS, Correia D, Prata A, Rodrigues Jr V. Familial analysis of seropositivity to *Trypanosoma cruzi* and of clinical forms of Chagas disease. *Am J Trop Med Hyg* 2010;**82**:45−8.

Downloaded From Vet-eBooks.com

7. Williams-Blangero S, VandeBerg JL, Blangero J, Correa-Oliveira R. Genetic epidemiology of Chagas disease. *Adv Parasitol* 2011;**75**:147−67.
8. Cunha-Neto E, Chevillard C. Chagas disease cardiomyopathy: immunopathology and genetics. *Mediat Inflamm* 2014:683230.
9. Ayo CM, Dalalio MMO, Visentainer JEL, Reis PG, Sippert EA, Jarduli LR, et al. Genetic susceptibility to Chagas disease: an overview about the infection and about the association between disease and the immune response genes. *Biomed Res Int* 2013:284729.
10. Shaw MA. Human genetic susceptibility to Chagas disease. In: Telleria J, Tibayrenc M, editors. *American trypanosomiasis: Chagas disease. One hundred years of research*. Elsevier; 2016 [in press].
11. Reis PG, Sell AM, Ayo CM, Oliveira CF, Dalalio MMO, Visentainer JV, et al. MHC class I polypeptide-related sequence A genes and linkage disequilibrium with HLA-B in Chagas disease. *Tissue Antigens* 2014;**84**:163.
12. Cruz-Robles D, Reyes PA, Monteon-Padilla VM, Ortiz-Muniz AR, Vargas-Alarcon G. MHC class I and class II genes in Mexican patients with Chagas disease. *Hum Immunol* 2004;**65**:60−5.
13. Deghaide NHS, Dantas RO, Donadi EA. HLA class I and II profiles of patients presenting with Chagas' disease. *Dig Dis Sci* 1998;**43**:246−52.
14. Layrisse Z, Fernandez MT, Montagnani S, Matos M, Balbas O, Herrera F, et al. HLA-C*03 is a risk factor for cardiomyopathy in Chagas disease. *Hum Immunol* 2000;**61**:925−9.
15. del Puerto F, Nishizawa JE, Kikuchi M, Roca Y, Avilas C, Gianella A, et al. Protective human leucocyte antigen haplotype, HLA-DRB1*01-B*14, against chronic Chagas disease in Bolivia. *PLoS Negl Trop Dis* 2012;**6**:e1587.
16. Lllop E, Rothhammer F, Acuna M, Apt W. HLA antigens in cardiomyopathic Chilean chagasics. *Am J Hum Genet* 1988;**43**:770−3.
17. Llop E, Rothhammer F, Acuna M, Apt W, Arribada A. HLA antigens in Chagasic heart-disease − new evidence based on a case control study. *Rev Med Chile* 1991;**119**:633−6.
18. Fernandez-Mestre MT, Layrisse Z, Montagnani S, Acquatella H, Catalioti F, Matos M, et al. Influence of the HLA class II polymorphism in chronic Chagas' disease. *Parasite Immunol* 1998;**20**:197−203.
19. Garcia Borras S, Racca L, Cotorruelo C, Biondi C, Beloscar J, Racca A. Distribution of HLA-DRB1 alleles in Argentinean patients with Chagas' disease cardiomyopathy. *Immunol Invest* 2009;**38**:268−75.
20. Borras SG, Diez C, Cotorruelo C, Pellizon O, Biondi C, Beloscar J, et al. HLA class II DRB1 polymorphism in Argentinians undergoing chronic *Trypanosoma cruzi* infection. *Ann Clin Biochem* 2006;**43**:214−6.
21. Nieto A, Beraun Y, Collado MD, Caballero A, Alonso A, Gonzalez A, et al. HLA haplotypes are associated with differential susceptibility to *Trypanosoma cruzi* infection. *Tissue Antigens* 2000;**55**:195−8.
22. Colorado IA, Acquatella H, Catalioti F, Fernandez MT, Layrisse Z. HLA class II DRB1, DQB1, DPB1 polymorphism and cardiomyopathy due to *Trypanosoma cruzi* chronic infection. *Hum Immunol* 2000;**61**:320−5.
23. Drigo SA, Cunha-Neto E, Ianni B, Cardoso MRA, Braga PE, Fae KC, et al. TNF gene polymorphisms are associated with reduced survival in severe Chagas' disease cardiomyopathy patients. *Microbes Infect* 2006;**8**:598−603.
24. Drigo SA, Cunha-Neto E, Ianni B, Mady C, Fae KC, Buck P, et al. Lack of association of tumor necrosis factor-alpha polymorphisms with Chagas disease in Brazilian patients. *Immunol Lett* 2007;**108**:109−11.

Downloaded From Vet-eBooks.com

25. Campelo V, Dantas RO, Simoes RT, Mendes-Junior CT, Sousa SMB, Simoes AL, et al. TNF microsatellite alleles in Brazilian Chagasic patients. *Dig Dis Sci* 2007;**52**:3334–9.
26. Criado L, Florez O, Martin J, Gonzalez CI. Genetic polymorphisms in *TNFA/TNFR2* genes and Chagas disease in a Colombian endemic population. *Cytokine* 2012;**57**:398–401.
27. Rodriguez-Perez JM, Cruz-Robles D, Hernandez-Pacheco G, Perez-Hernandez N, Murguia LE, Granados J, et al. Tumor necrosis factor-alpha promoter polymorphism in Mexican patients with Chagas' disease. *Immunol Lett* 2005;**98**:97–102.
28. Alves SM, Lannes Vieira J, Arnez LEA, Moraes MO, Oliveira WA, Sarteschi C, et al. Chagas cardiomyopathy: prognostic value of genetic polymorphisms of TNF-alpha. *Eur J Heart Fail* 2014;**16**:250.
29. Pissetti CW, Correia D, de Oliveira RF, Llaguno MM, Balarin MAS, Silva-Grecco RL, et al. Genetic and functional role of TNF-alpha in the development *Trypanosoma cruzi* infection. *PLoS Negl Trop Dis* 2011;**5**:e976.
30. Beraun Y, Nieto A, Collado MD, Gonzalez A, Martin J. Polymorphisms at tumor necrosis factor (TNF) loci are not associated with Chagas' disease. *Tissue Antigens* 1998;**52**:81–3.
31. Pissetti CW, de Oliveira RF, Correia D, Nascentes GAN, Llaguno MM, Rodrigues Jr V. Association between the lymphotoxin-alpha gene polymorphism and Chagasic cardiopathy. *J Interferon Cytokine Res* 2013;**33**:130–5.
32. Ramasawmy R, Fae KC, Cunha-Neto E, Muller NG, Cavalcanti VL, Ferreira RC, et al. Polymorphisms in the gene for lymphotoxin-alpha predispose to chronic Chagas cardiomyopathy. *J Infect Dis* 2007;**196**:1836–43.
33. del Puerto F, Kikuchi M, Nishizawa JE, Roca Y, Avila C, Gianella A, et al. 21-Hydroxylase gene mutant allele CYP21A2*15 strongly linked to the resistant HLA haplotype B*14:02-DRB1*01:02 in chronic Chagas disease. *Hum Immunol* 2013;**74**:783–6.
34. Arnez LEA, Venegas EN, Ober C, Thompson EE. Sequence variation in the *IL4* gene and resistance to *Trypanosoma cruzi* infection in Bolivians. *J Allergy Clin Immunol* 2011;**127**: 279–82.
35. Florez O, Zafra G, Morillo C, Martin J, Gonzalez CI. Interleukin-1 gene cluster polymorphism in Chagas disease in a Colombian case-control study. *Hum Immunol* 2006;**67**: 741–8.
36. Cruz-Robles D, Pablo Chavez-Gonzalez J, Magdalena Cavazos-Quero M, Perez-Mendez O, Reyes PA, Vargas-Alarcon G. Association between IL-1B and IL-1RN gene polymorphisms and Chagas' disease development susceptibility. *Immunol Invest* 2009;**38**: 231–9.
37. Rodriguez DAL, Carmona FD, Echeverria LE, Gonzalez CI, Martin J. *IL18* gene variants influence the susceptibility to Chagas disease. *PLoS Negl Trop Dis* 2016;**10**:e0004583.
38. Nogueira LG, Frade AF, Ianni BM, Laugier L, Pissetti CW, Cabantous S, et al. Functional IL18 polymorphism and susceptibility to chronic Chagas disease. *Cytokine* 2015;**73**:79–83.
39. Torres OA, Calzada JE, Beraun Y, Morillo CA, Gonzalez A, Gonzalez CI, et al. Role of the IFNG+874T/A polymorphism in Chagas disease in a Colombian population. *Infect Genet Evol* 2010;**10**:682–5.
40. Leon Rodriguez DA, Echeverria LE, Gonzalez CI, Martin J. Investigation of the role of IL17A gene variants in Chagas disease. *Genes Immun* 2015;**16**:536–40.
41. Calzada JE, Beraun Y, Gonzalez CI, Martin J. Transforming growth factor beta 1 (TGF beta 1) gene polymorphisms and Chagas disease susceptibility in Peruvian and Colombian patients. *Cytokine* 2009;**45**:149–53.
42. Florez O, Martin J, Gonzalez CI. Interleukin 4, interleukin 4 receptor-alpha and interleukin 10 gene polymorphisms in Chagas disease. *Parasite Immunol* 2011;**33**:506–11.

Downloaded From Vet-eBooks.com

43. Zafra G, Morillo C, Martin J, Gonzalez A, Gonzalez CI. Polymorphism in the 3′ UTR of the *IL12B* gene is associated with Chagas' disease cardiomyopathy. *Microbes Infect* 2007;**9**: 1049−52.
44. de Oliveira AP, Bernardo CR, Camargo AV, Ronchi LS, Borim AA, Brandao de Mattos CC, et al. Genetic susceptibility to cardiac and digestive clinical forms of chronic Chagas disease: involvement of the CCR5 59029 A/G polymorphism. *PLoS One* 2015;**10**: e0141847.
45. Calzada JE, Nieto A, Beraun Y, Martin J. Chemokine receptor CCR5 polymorphisms and Chagas' disease cardiomyopathy. *Tissue Antigens* 2001;**58**:154−8.
46. Fernandez-Mestre MT, Montagnani S, Layrisse Z. Is the CCR5-59029-G/G genotype a protective factor for cardiomyopathy in Chagas disease? *Hum Immunol* 2004;**65**:725−8.
47. Machuca MA, Suarez EU, Echeverria LE, Martin J, Gonzalez CI. SNP/haplotype associations of CCR2 and CCR5 genes with severity of chagasic cardiomyopathy. *Hum Immunol* 2014;**75**:1210−5.
48. Florez O, Martin J, Gonzalez CI. Genetic variants in the chemokines and chemokine receptors in Chagas disease. *Hum Immunol* 2012;**73**:852−8.
49. Frade AF, Pissetti CW, Ianni BM, Saba B, Wang HTL, Nogueira LG, et al. Genetic susceptibility to Chagas disease cardiomyopathy: involvement of several genes of the innate immunity and chemokine-dependent migration pathways. *BMC Infect Dis* 2013;**13**: 587.
50. Nogueira LG, Barros Santos RH, Ianni BM, Fiorelli AI, Mairena EC, Benvenuti LA, et al. Myocardial chemokine expression and intensity of myocarditis in Chagas cardiomyopathy are controlled by polymorphisms in *CXCL9* and *CXCL10*. *PLoS Negl Trop Dis* 2012; **6**:e1867.
51. Ramasawmy R, Cunha-Neto E, Fae KC, Martello FG, Muller NG, Cavalcanto VL, et al. The monocyte chemoattractant protein-1 gene polymorphism is associated with cardiomyopathy in human Chagas disease. *Clin Infect Dis* 2006;**43**:305−11.
52. Weitzel T, Zulantay I, Danquah I, Hamann L, Schumann RR, Apt W, et al. Short report: Mannose-binding lectin and toll-like receptor polymorphisms and Chagas disease in Chile. *Am J Trop Med Hyg* 2012;**86**:229−32.
53. Ramasawmy R, Cunha-Neto E, Fae KC, Borba SCP, Teixeira PC, Ferreira SCP, et al. Heterozygosity for the S180L variant of *MAL/TIRAP*, a gene expressing an adaptor protein in the Toll-like receptor pathway, is associated with lower risk of developing chronic Chagas cardiomyopathy. *J Infect Dis* 2009;**199**:1838−45.
54. Zafra G, Florez O, Morillo CA, Echeverria LE, Martin J, Gonzalez CI. Polymorphisms of toll-like receptor 2 and 4 genes in Chagas disease. *Mem Inst Oswaldo Cruz* 2008;**103**: 27−30.
55. Luz PR, Miyazaki MI, Neto NC, Padeski MC, Barros ACM, Boldt ABW, et al. Genetically determined MBL deficiency is associated with protection against chronic cardiomyopathy in Chagas disease. *PLoS Negl Trop Dis* 2016;**10**:e0004257.
56. Boldt ABW, Luz PR, Messias-Reason IJT. MASP2 haplotypes are associated with high risk of cardiomyopathy in chronic Chagas disease. *Clin Immunol* 2011;**140**:63−70.
57. Luz PR, Boldt ABW, Grisbach C, Kun JFJ, Velavan TP, Messias-Reason IJT. Association of L-ficolin levels and FCN2 genotypes with chronic Chagas disease. *PLoS One* 2013;**8**: e60237.
58. Dias FC, Medina TS, Mendes-Junior CT, Dantas RO, Pissetti CW, Rodrigues Jr V, et al. Polymorphic sites at the immunoregulatory CTLA-4 gene are associated with chronic Chagas disease and its clinical manifestations. *PLoS One* 2013;**8**:e78367.

Downloaded From Vet-eBooks.com

59. Fernandez-Mestre M, Sanchez K, Balbas O, Gendzekhzadze K, Ogando V, Cabrera M, et al. Influence of CTLA-4 gene polymorphism in autoimmune and infectious diseases. *Hum Immunol* 2009;**70**:532–5.
60. Jorge SEDC, Abreu CF, Guariento ME, Sonati MF. Haptoglobin genotypes in Chagas' disease. *Clin Biochem* 2010;**43**:314–6.
61. Fernandez NM, Fernandez-Mestre M. The role of haptoglobin genotypes in Chagas disease. *Dis Markers* 2014:793646.
62. Bush WS, Moore JH. Chapter 11: Genome-wide association studies. *PLoS Comput Biol* 2012;**8**:e1002822.
63. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet* 2012;**90**:7–24.
64. Deng X, Sabino EC, Cunha-Neto E, Ribeiro AL, Ianni B, Mady C, et al. Genome wide association study (GWAS) of Chagas cardiomyopathy in *Trypanosoma cruzi* seropositive subjects. *PLoS One* 2013;**8**:e79629.
65. Theis JH, Tibayrenc M, Mason DT, Ault SK. Exotic stock of *Trypanosoma cruzi* (*Schizotrypanum*) capable of development in and transmission by *Triatoma protracta protracta* from California. Public health implications. *Am J Trop Med Hyg* 1987;**36**:523–8.
66. Tibayrenc M. Modeling the transmission of *Trypanosoma cruzi*: the need for an integrated genetic epidemiological and population genomics approach. In: Landes ME, editor. *Infectious disease transmission modeling and management of parasite control*. Bioscience/Eurekah; 2009.
67. Toyé PJ. Isoenzymic differences between culture forms of *Trypanosoma rangeli*, *T. cruzi*, and *T. lewisi*. *Trans R Soc Trop Med Hyg* 1974;**68**:266.
68. Miles MA, Souza A, Povoa M, Shaw JJ, Lainson R, Toyé PJ. Isozymic heterogeneity of *Trypanosoma cruzi* in the first autochtonous patients with Chagas' disease in Amazonian Brazil. *Nature* 1978;**272**:819–21.
69. Tibayrenc M, Ward P, Moya A, Ayala FJ. Natural populations of *Trypanosoma cruzi*, the agent of Chagas' disease, have a complex multiclonal structure. *Proc Nat Acad Sci USA* 1986;**83**:115–9.
70. Tibayrenc M, Kjellberg F, Ayala FJ. A clonal theory of parasitic protozoa: the population structure of *Entamoeba, Giardia, Leishmania, Naegleria, Plasmodium, Trichomonas* and *Trypanosoma*, and its medical and taxonomical consequences. *Proc Nat Acad Sci USA* 1990;**87**:2414–8.
71. Tibayrenc M, Kjellberg F, Arnaud J, Oury B, Brenière SF, Dardé ML, et al. Are eukaryotic microorganisms clonal or sexual? A population genetics vantage. *Proc Natl Acad Sci USA* 1991;**88**:5129–33.
72. Tibayrenc M, Ayala FJ. Reproductive clonality of pathogens: a perspective on pathogenic viruses, bacteria, fungi, and parasitic protozoa. *Proc Nat Acad Sci USA* 2012;**109**(48):E3305–13.
73. Rougeron V, De Meeûs T, Hide M, Waleckx E, Bermudez H, Arevalo J, et al. Extreme inbreeding in *Leishmania braziliensis*. *Proc Nat Acad Sci USA* 2009;**25**:10224–9.
74. Rougeron V, De Meeûs T, Kako Ouraga S, Hide M, Bañuls AL. ''Everything you always wanted to know about sex (but were afraid to ask)'' in *Leishmania* after two decades of laboratory and field analyses. *PLoS Pathog* 2010;**6**(8):e1001004. http://dx.doi.org/10.1371/journal.ppat.1001004.
75. Machado CA, Ayala FJ. Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*. *Proc Nat Acad Sci USA* 2001;**98**:7396–401.

Downloaded From Vet-eBooks.com

76. Gaunt MW, Yeo M, Frame IA, Tothard JR, Carrasco HJ, Taylor MC, et al. Mechanism of genetic exchange in American trypanosomes. *Nature* 2003;**421**:936−9.
77. Tibayrenc M, Neubauer K, Barnabé C, Guerrini F, Sarkeski D, Ayala FJ. Genetic characterization of six parasitic protozoa: parity of random-primer DNA typing and multilocus isoenzyme electrophoresis. *Proc Natl Acad Sci USA* 1993;**90**:1335−9.
78. Barnabé C, Brisse S, Tibayrenc M. Population structure and genetic typing of *Trypanosoma cruzi*, the agent of Chagas' disease: a multilocus enzyme electrophoresis approach. *Parasitol* 2000;**150**:513−26.
79. Brisse S, Barnabé C, Tibayrenc M. Identification of six *Trypanosoma cruzi* phylogenetic lineages by random amplified polymorphic DNA and multilocus enzyme electrophoresis. *Int J Parasitol* 2000;**30**:35−44.
80. Brisse S, Henriksson J, Barnabé C, Douzery EJP, Berkvens D, Serrano M, et al. Evidence for genetic exchange and hybridization in *Trypanosoma cruzi* based on nucleotide sequences and molecular karyotype. *Infect Genet Evol* 2003;**2**:173−83.
81. Sturm N, Campbell DA. Alternative lifestyles: the population structure of *Trypanosoma cruzi*. *Acta Trop* 2010;**115**:35−43.
82. Zingales B, Andrade SG, Briones MRS, Campbell DA, Chiari E, Fernandes O, et al. A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. *Mem Inst Oswaldo Cruz* 2009;**104**:1051−4.
83. Zingales B, Miles MA, Campbell D, Tibayrenc M, Macedo AM, Teixeira MM, et al. The revised *Trypanosoma cruzi* subspecific nomenclature: rationale, epidemiological relevance and research applications. *Infect Genet Evol* 2012;**12**:240−53.
84. Avise JC. *Molecular markers, natural history and evolution*. 2nd ed. New York, London: Chapman and Hall; 2004.
85. Tibayrenc M, Ayala FJ. The population genetics of *Trypanosoma cruzi* revisited in the light of the predominant clonal evolution model. *Acta Trop* 2015;**151**:156−65.
86. Messenger LA, Llewellyn MS, Bhattacharyya T, Franzén O, Lewis MD, Ramírez JD, et al. Multiple mitochondrial introgression events and heteroplasmy in *Trypanosoma cruzi* revealed by maxicircle MLST and next generation sequencing. *PLoS Negl Trop Dis* 2012;**6**: e1584. http://dx.doi.org/10.1371/journal.pntd.0001584.
87. Ramírez JD, Llewellyn MS. Reproductive clonality in protozoan pathogens—truth or artefact? *Mol Ecol* 2014;**23**:4195−202.
88. Ramírez JD, Llewellyn MS. Response to Tibayrenc and Ayala: reproductive clonality in protozoan pathogens − truth or artefact? *Mol Ecol* 2015;**24**:5782−4.
89. Lima L, Espinosa-Álvarez O, Ortiz PA, Trejo-Varón JA, Carranza JC, Pinto CM, et al. Genetic diversity of *Trypanosoma cruzi* in bats, and multilocus phylogenetic and phylogeographical analyses supporting Tcbat as an independent DTU (discrete typing unit). *Acta Trop* 2015;**151**:166−77.
90. Marcili A, Lima L, Cavazzsana M, Junqueira ACV, Veludo HH, Maia da Silva F, et al. A new genotype of *Trypanosoma cruzi* associated with bats evidenced by phylogenetic analyses using SSU rDNA, cytochrome b and Histone H2B genes and genotyping based on ITS1 rDNA. *Parasitol* 2009;**136**:641−55.
91. Pinto CM, Kalko EKV, Cottontail I, Wellinghausen N, Cottontail VM. TcBat a bat-exclusive lineage of *Trypanosoma cruzi* in the Panama Canal Zone, with comments on its classification and the use of the 18S rRNA gene for lineage identification. *Infect Genet Evol* 2012;**12**:1328−32.
92. Pinto CM, Ocaña-Mayorga S, Tapia EE, Lobos SE, Zurita AP, Aguirre-Villacís F, et al. Bats, trypanosomes, and triatomines in Ecuador: new insights into the diversity, transmission, and origins of *Trypanosoma cruzi* and Chagas disease. *PLoS One* 2016;**10**(10): e0139999. http://dx.doi.org/10.1371/journal.pone.0139999.

Downloaded From Vet-eBooks.com

93. Barnabé C, Mobarec HI, Jurado MR, Cortez JA, Brenière SF. Reconsideration of the seven discrete typing units within the species *Trypanosoma cruzi*, a new proposal of three reliable mitochondrial clades. *Infect Genet Evol* 2016;**39**:176−86.
94. Tibayrenc M, Ayala FJ. How clonal are *Trypanosoma* and *Leishmania*? *Trends Parasitol* 2013;**29**:264−9.
95. Revollo S, Oury B, Laurent JP, Barnabé C, Quesney V, Carrière V, et al. *Trypanosoma cruzi*: impact of clonal evolution of the parasite on its biological and medical properties. *Exp Parasitol* 1998;**89**:30−9.
96. Telleria J, Biron DG, Brizard JP, Demettre E, Seveno M, Barnabé C, et al. Phylogenetic character mapping of proteomic diversity shows high correlation with subspecific phylogenetic diversity in *Trypanosoma cruzi*, the agent of Chagas disease. *Proc Nat Acad Sci USA* 2010;**107**:20411−6.
97. Machin A, Telleria J, Brizard JP, Demettre E, Séveno M, Ayala FJ, et al. *Trypanosoma cruzi*: gene expression surveyed by proteomic analysis reveals interaction between different genotypes in mixed in vitro cultures. *PLoS One* 2014;**9**(4). ISSN: 1932-6203:e95442.
98. Tibayrenc M. Integrated genetic epidemiology of infectious diseases: the Chagas model. *Mem Inst Oswaldo Cruz* 1998;**93**:577−80.

Downloaded From Vet-eBooks.com

Adaptive Evolution of the *Mycobacterium tuberculosis* Complex to Different Hosts

23

E. Broset [1,2], *J. Gonzalo-Asensio* [1,2,3]
[1]Universidad de Zaragoza, Zaragoza, Spain; [2]CIBERes, Instituto de Salud Carlos III, Madrid, Spain; [3]Hospital Universitario Miguel Servet, Zaragoza, Spain

1. Overview: Disease and Mycobacterial Genetics

The *Mycobacterium* genus belongs to the phylum Actinobacteria, and it comprises Gram-positive species with a high genomic content (GC) (ranging from 69% to 58%). In contrast to other Gram-positive bacteria, mycobacteria possess a multilayered cell envelope rich in uncommon lipids responsible for their distinctive Zielh–Neelsen staining. Mycobacteria are widely distributed in either the environment or infecting different hosts and they are subdivided into rapid- and slow-growing species, based on their ability to develop colonies in less or more than 7 days, respectively. Fast-growing species are in general opportunistic or nonpathogenic bacteria, whereas slow growers include major human pathogenic mycobacteria, such as *Mycobacterium tuberculosis*, *Mycobacterium ulcerans*, or *Mycobacterium leprae* causing tuberculosis (TB), Buruli ulcer or leprosy, respectively. This chapter focuses on the *M. tuberculosis* complex (MTBC) comprising a group of closely related subspecies or ecotypes adapted to cause tuberculosis (TB) in different mammalian hosts including humans.

To better understand mycobacterial genetics, we briefly describe the TB infectious cycle. TB in humans is not restricted to the lungs since it may affect any organ. Also, as is detailed in the following paragraphs, other hosts aside from humans are also susceptible to TB. Since our clinical and scientific knowledge comes primarily from human samples, we will predominantly refer to human pulmonary TB, albeit readers must be conscious that other manifestations exist. The infectious cycle starts transmitting the tubercle bacilli by respiratory route. When a patient with active pulmonary disease coughs, sneezes, or even speaks, the aerosolized bacteria are then inhaled by neighboring individuals and they reach the alveoli where resident macrophages phagocyte them. The fate of most bacteria in the phagosome is often fatal but *M. tuberculosis* is able to survive—and even to escape—from the phagosome. Once engulfed, *M. tuberculosis* triggers a host–pathogen cross talk causing a specific activation of the host immune system resulting in a nucleus of infected macrophages surrounded by a mantle of activated T cells and enclosed by a fibrous cuff. Within this structure called granuloma, *M. tuberculosis* can survive for decades in an asymptomatic state and it is estimated that one-third of the human population are latent carriers of the

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00023-8
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

tubercle bacilli. In order to transmit from person to person, this containment must fail, which usually happens as a consequence of immune system weakening. Then, the granuloma breaks down causing a lesion in the lung where *M. tuberculosis* multiplies extracellularly, reaching the airways where bacilli can spread and the infectious cycle continues in a new infected person. The maintenance of *M. tuberculosis* in the host population relies on the ability to tightly regulate rounds of infection, reactivation to disease, and transmission. As an illustrative example, disease occurs in about 5−10% of the infected individuals, which likely represent the optimal percentage to ensure *M. tuberculosis* survival in the human population. It is possible that higher disease rates would decimate susceptible hosts, while lower rates would not ensure proper transmission rates. As is discussed in this chapter, TB-causing bacteria have likely adapted their virulence and consequently their transmissibility depending on the population density. Unlike other microorganisms that rely on antigenic diversity to escape from the immune system, MTBC members have emerged from a genetic bottleneck and consequently show little DNA diversity. This is particularly well applied to the repertoire on human T-cell epitopes, which are hyperconserved among *M. tuberculosis* strains and point at this bacterium as a professional pathogen evolved for being recognized by the host immune system.

Our genetic knowledge of TB-causing bacteria required the development of suitable genetic tools (plasmids, phages, transposons, and gene replacement systems). These pioneering studies were initiated by two independent laboratories led by Williams R. Jacobs at Albert Einstein Institute (New York)[1] and Brigitte Gicquel at Pasteur Institute (Paris).[2] Subsequent genetic studies sought to restore virulence to an *M. tuberculosis* attenuated strain—namely H37Ra—through genomic complementation with DNA from its virulent counterpart—namely H37Rv—resulting in the identification of a genomic segment with a potential role in virulence.[3] Another founding study identified the IS*6110* transposon in *M. tuberculosis* and demonstrated its utility as epidemiological marker based on its exclusive presence in MTBC members and the variation in copy number and location of this transposon between different strains.[4] This study paved the way for "modern" molecular epidemiology in TB and today, 25 years after its original description, IS*6110* is worldwide used in *M. tuberculosis* epidemiology.[5]

Pregenomic studies used subtractive hybridization of an "interrogated" genome to a reference genome to identify large deletions in the "interrogated" DNA. Using this approach, several deletions (designated Regions of Difference, RD1 to RD3) were identified in the BCG vaccine relative to its parental *Mycobacterium bovis* strain or to *M. tuberculosis*.[6,7] RD1 was latter demonstrated to be the main contributor to BCG vaccine attenuation. A genome map of the *M. tuberculosis* H37Rv reference strain was delineated using a combination of pulse field electrophoresis and hybridization with cosmid libraries. This study also allowed the identification of "hot spots" sites for IS*6110* transposition and a genomic comparison with *M. leprae*.[8] Subsequent studies used bacterial artificial chromosomes (BACs) to refine genome maps. Further, since this BAC library achieved almost total coverage of the *M. tuberculosis* chromosome, it was a powerful tool for the H37Rv genome-sequencing project.[9] BAC libraries also allowed to gain resolution in subtractive hybridization experiments to

Downloaded From Vet-eBooks.com

identify seven novel RD deletions (RD4 to RD10) in *M. bovis* relative to H37Rv. Some RD deletions were found in *Mycobacterium africanum* (RD7 to RD10) and *M. bovis* (RD4) and this pioneering study defined the evolutive steps suffered by MTBC members to adapt to specific hosts.[10]

More than a century after the initial description of *M. tuberculosis* by Robert Koch, the complete genome sequence of the *M. tuberculosis* H37Rv strain was deciphered. It consists of a 4.4 Mb genome coding for 3959 ORF, 45 tRNA, and 1 rRNA operon. The 65% G + C content impacts on the biased amino acid composition. An important proportion of the *M. tuberculosis* genome is devoted to lipogenesis and lipolysis, which reflects the particular lifestyle of the bacterium. Another finding from the genome sequence is the high content of proteins with repetitive structures whose role has not yet been elucidated.[11]

2. Host–Pathogen Coevolution of the Tubercle Bacillus

Publication of the genome sequence of *M. tuberculosis* H37Rv[11] was a scientific landmark to start deciphering the biology of *M. tuberculosis* since it opened new perspectives and enabled "omic" approaches. Development of H37Rv DNA probes for microarrays were used to either compare transcriptomes or to analyze genomes from related mycobacteria by comparative genomics. Identification of variable regions in *Mycobacterium* species and the finding that these genomic deletions does not occur independently in the different strains of the MTBC, led to propose an evolutionary landscape for the MTBC. This phylogeny is based on the idea of irreversible loss of genetic regions that cannot be compensated by horizontal gene transfer (HGT). The pattern of reductive evolution is associated with an intracellular lifestyle (which is well reflected in the obligate pathogen *M. leprae* possessing a highly reduced genome) and also consistent with the presence of a genetic bottleneck at the split between MTBC and the progenitor population.

According to distribution of RD deletions, MTBC species can be classified into human-adapted (*M. tuberculosis*, *M. africanum*) and animal-adapted species (*M. bovis* and related ecotypes).[12] This evolutionary scenario allowed to revisit the pre-existing hypothesis that human-infecting mycobacteria arose through zoonotic transmission of *M. bovis* during cattle domestication. In contrast, this phylogeny revealed that *M. bovis* (and animal-adapted strains) have accumulated deletions with respect to human-adapted lineages and consequently derive from these latter. Archaeological records also conflict with the zoonotic hypothesis since evidences for human TB have been found in human remains dating prior to animal domestication (Fig. 23.1A and D).

Analysis of whole genomes from 259 MTBC strains has served to establish a genome-based phylogeny that is congruent with deletion-based phylogeny. MTBC comprises eight major lineages (L1 to L8) which include the human-adapted ecotypes *M. tuberculosis* (L1 to L4 and L7), *M. africanum* (L5 and L6) and the animal-adapted ecotypes *M. bovis*, *Mycobacterium caprae*, *Mycobacterium microti*, *Mycobacterium*

Downloaded From Vet-eBooks.com

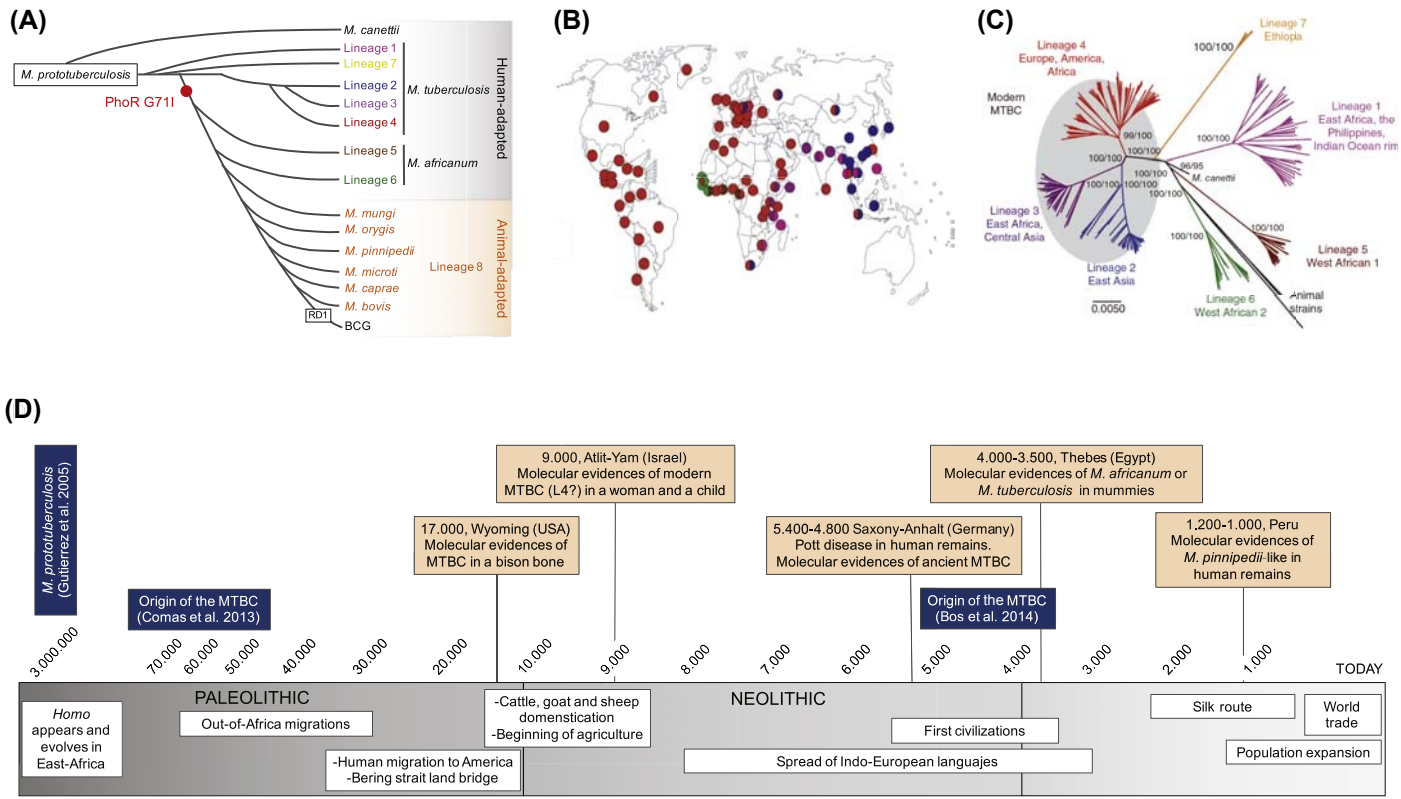


Figure 23.1 (A) Deletion-based phylogeny of the MTBC showing the human- and animal-adapted lineages. With the exception of RD1 and Gly71Ile in *phoR*, the remaining deletions and polymorphisms have been omitted for clarity. Please note that *Mycobacterium canettii* does not belong to the MTBC. (B) Phylogeographical distribution on the main MTBC lineages causing TB in humans. L2, L3, and L4 strains are globally distributed, while

Downloaded From Vet-eBooks.com

L1, L5, L6, and L7 lineages are geographically restricted. (C) Polymorphism-based phylogeny of the MTBC. Note the similarity with deletion-based phylogeny. (D) Timeline showing the main historic events of humankind in relation to MTBC evolution. Orange boxes (*grey in print version*) indicate relevant archaeological and molecular evidences for MTBC in human and animal remains. Dark blue boxes (*dark grey in print version*) indicate estimates about the origin of *Mycobacterium prototuberculosis* and the MTBC.
Images have been adapted from Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, et al. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat Genet* 2013;**45**(10):1176−82; Comas I, Chakravartti J, Small PM, Galagan J, Niemann S, Kremer K, et al. Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat Genet* 2010;**42**(6): 498−503; Broset E, Martin C, Gonzalo-Asensio J. Evolutionary landscape of the *Mycobacterium tuberculosis* complex from the Viewpoint of PhoPR: implications for virulence regulation and application to vaccine development. *MBio* 2015;**6**(5).

Downloaded From Vet-eBooks.com

pinnipedii, *Mycobacterium orygis*, and *Mycobacterium mungi* grouped into L8 (Fig. 23.1C). Similar to L8 species that have evolved to infect specific mammals, human-adapted strains have evolved to cause TB in human subpopulations. Accordingly, *Mycobacterium canettii* was originally isolated in the horn of Africa,[13] *M. africanum* L5 and L6 strains are commonly found in west African countries,[14] *M. tuberculosis* L7 is frequent in Ethiopia, L1 in the Indian Ocean rim and L2, L3, and L4, which show wider distribution, predominantly infect people from east Asia, east India, and America−Europe, respectively[15] (Fig. 23.1B). A detailed look at the evolution of the Beijing lineage (L2), which is associated with the massive spread of drug resistance in Eurasia, suggest that worldwide colonization with this lineage started with the onset of agriculture in China, Korea, and Japan about 6000 years ago. Drug resistance associated with L2 strains appears to be related with positive selection of single-nucleotide polymorphisms (SNPs) in selected genes.[16]

Phylogenies of the MTBC are useful to estimate the most recent common ancestor (MRCA) of the tubercle bacillus. *Mycobacterium canettii* (or *Mycobacterium prototuberculosis*) predated the clonal expansion of the MTBC, and therefore, this species is considered the progenitor of the tubercle bacillus. The origin of *M. canettii* is situated 2.6−2.8 million years ago. Thus, TB is older than plague, typhoid fever, or malaria and might have affected early hominids. Consistent with this hypothesis, Australopithecus and *Homo habilis* were present 3 million years ago in the horn of Africa, a region where most *M. canettii* strains have been isolated[17] (Fig. 23.1D). *M. canettii* presents particular genomic characteristics with respect to MTBC members: (1) there are evidences for HGT and recombination events, (2) it presents larger genome sizes in contrast with the reductive evolution of the MTBC, and (3) there is a 25-fold higher frequency of SNPs compared to MTBC. These features suggest that *M. canettii* diverged from the MRCA of the tubercle bacilli before the clonal expansion of the MTBC.[18]

Assuming a constant mutation rate, it has been possible to date the origin of the MTBC speciation. However, there is still debate about the exact date when speciation occurred. Some studies situate this origin less than 6000 years ago,[19] while others propose a more ancient origin around 70,000 years ago.[15] The oldest archaeological evidence of MTBC comes from a bison bone dated back 17,000 years ago and molecular clocks should predate this irrefutable proof. Domestication of cattle began 10,500 years ago and molecular evidences for the presence of MTBC DNA in this ancient bison reinforce the idea that even if TB arose as a zoonosis, it predated animal domestication (Fig. 23.1D). Considering the ancient origin and genomic features of *M. canettii*, the genetic bottleneck of the MTBC prior to animal domestication and the lack of evidence for human-to-human transmission of *M. canettii*, it has been proposed an opportunistic infection from an environmental reservoir for the progenitor of the tubercle bacilli. Eventually, this opportunistic bacterium would have evolved as a professional pathogen in mammalian hosts.[20]

On another scenario, assuming that *M. bovis* and other animal-adapted strains derive from *M. tuberculosis*, this implies that the MRCA was adapted to humans and consequently humans would have given TB to animals via "reverse" zoonosis.[21] Supporting this hypothesis, since *M. africanum* lies somewhere between *M. tuberculosis* and animal-adapted species in the phylogenetic tree, together with the high prevalence of

Downloaded From Vet-eBooks.com

M. africanum in West African countries and the sporadic isolation of *M. africanum* from monkeys and cows in this region,[14] we can hypothesize that these isolates represent an opportunistic spill over transmission into animals from humans.

Association of MTBC members with specific mammals represent an attractive example of host−pathogen adaptation. Considering the highly conserved genome of MTBC species, it is intriguing to think that two humans chosen at random have twice (0.1%) the sequence diversity than human-adapted *M. tuberculosis* and cattle-adapted *M. bovis* (0.05%). However, after a critical examination of these numbers, a 0.05% sequence divergence is translated into about 2000 polymorphisms between two MTBC strains. Considering that these species code for roughly 4000 genes, it is inferred that half of the genes are virtually affected by polymorphisms. Even if some polymorphisms represent silent mutations, it is tempting to assume that some polymorphisms might have determined the host range of MTBC. As detailed in Section 6, the profound implications in virulence for a single amino acid mutation in *M. bovis* and *M. africanum* with respect to *M. tuberculosis* strains[22] was recently demonstrated (Fig. 23.1A).

Hereafter, we focus on the host−pathogen evolution for the most relevant MTBC species that cause TB in humans and other mammals.

2.1 Mycobacterium tuberculosis

This bacterium is the major, but not the unique, cause of TB in humans. TB caused 20% mortality between the 17th and 19th centuries and remains today the most lethal infectious disease. The close association of *M. tuberculosis* with humans is demonstrated by the striking similarity of the phylogenetic trees from 220 strains of the MTBC and 4.955 mitochondrial genomes representative of the main human haplogroups. The incidence of TB was not only markedly increased as a consequence of the Neolithic expansion but also this period triggered an increment in the population diversity of the MTBC and humans. Therefore, it is suggested that changes in human demography have shaped the evolution of *M. tuberculosis* as demonstrated by the existence of different lineages with specific geographical distribution. Further, dating estimates of the different lineages remarkably correlate with human migratory movements. Dispersal of modern humans occurred in two major waves: an initial dispersal around the Indian Ocean during 62,000−75,000 years ago and a later dispersal into Eurasia 25,000−38,000 years ago. The first wave would have given rise to L1 (East Africa, Philippines, and Indian Ocean rim) and the second wave to L2 (East Asia), L3 (East Africa, Central Asia), and L4 (Europe, America). Later migrations, conquests, and slave trade would have contributed to the worldwide spread of L4 strains into America and Africa. The L7 (Ethiopia) might have arisen among a human population that remained in or returned to Africa. The African origin of modern humans and the coalescence of all MTBC branches in this continent reinforces the hypothesis of the human origin of TB.[15,20]

2.2 Mycobacterium africanum

The two lineages of *M. africanum* (L5 and L6) constitute the second major cause of human TB (Fig. 23.1B). Although *M. africanum* has been identified alongside

Downloaded From Vet-eBooks.com

M. tuberculosis in human remains from Egypt dating back 4000 years ago, today *M. africanum* is geographically restricted to West Africa where they show a prevalence ranging to 21−66%. Out-of-Africa cases of *M. africanum*-related TB are rare and frequently linked to African immigrants.[14] This highlights the exquisite adaptation of this species to a geographically restricted population. It has been reported that although *M. tuberculosis* and *M. africanum* L6 are equally transmitted in the Gambia, a 2-year follow-up demonstrated a fivefold reduction in *M. africanum* L6 relative to *M. tuberculosis* in progression to active disease.[23] It is also remarkable that *M. africanum* did not established in the New World despite massive slave trade from 15th to 19th centuries, which likely indicates that *M. africanum* was readily outcompeted by *M. tuberculosis* in the New World. Further, the phylogenetic relatedness of *M. africanum* with animal strains, the evidence of *M. africanum*-infected animals and the low transmission to disease in humans might point at this specie either as a spill over transmission to animals by reverse zoonosis or an animal reservoir for *M. africanum*.

2.3 Mycobacterium canettii

It is important to note that *M. canettii* is considered an ancestral lineage from which the MTBC emerged and consequently does not belong to this latter. Alongside *M. tuberculosis* and *M. africanum*, *M. canettii* cause human TB. In contrast, *M. canettii* exhibit distinctive smooth colony morphology in culture media and human isolates are geographically restricted to East African populations. It is estimated that this species emerged 2.8 million years ago, which mean that early hominids in East Africa would have been infected with a TB-like disease (Fig. 23.1D).[17] At the genome level, *M. canettii* strains have distinctive features compared with the MTBC. Whole-genome sequencing of five representative isolates from different subtypes (A, D, L, J, and K) showed a similar organization and synteny with MTBC genomes. However, *M. canettii* genomes are 10−115 kb larger than those of the MTBC and they present evidences for HGT and recombination that resulted in a genetic mosaicism of *M. canettii* inherited by MTBC. Phenotypically, *M. canettii* strains grow faster than *M. tuberculosis* especially at low temperatures (30°C) and show reduced persistence in the mouse model of aerosol infection. These features reflecting broader environmental adaptability, but lower virulence and transmission makes us to think on *M. canettii* as a missing link between the obligate pathogens of the MTBC and a putative progenitor environmental mycobacteria.[18] Genomic comparison of *M. canettii* strains belonging to a pathogenic outbreak, has raised the question about the eventual increased virulence of particular strains. This study suggests that current epidemic clones of *M. canettii* might mimic an earlier phase before MTBC speciation. Thus, the eventual acquisition of the appropriate mutations, either in the environmental reservoir or within an infected host, would enable person to person transmission of *M. canettii*.[24] In this line, a study has described a role for lipooligosaccharide (LOS) synthesis in host−pathogen interaction and in *M. canettii* virulence. Smooth

Downloaded From Vet-eBooks.com

M. canetii variants producing LOS are less virulent than rough variants lacking LOS. Further insight into the biological mechanism indicates that recombination between two *pks5* genes is responsible for this loss of LOS. Altogether this recombination-mediated surface remodeling might represent an evolutionary step to shift from the putative generalist *M. canettii* to the professional pathogens of the MTBC.[25]

2.4 Mycobacterium bovis

TB in animals is mainly known from cases in cattle caused by *M. bovis* and referred to as bovine TB. *M. bovis* is 99.95% identical to *M. tuberculosis*, the about 2000 polymorphisms, insertions, and deletions present between both strains clearly bias the host tropism. *M. bovis* requires pyruvate when glycerol is the primary carbon source due to an inactive pyruvate kinase. This finding in addition to have implications during laboratory work will surely have functional implications in the infection of the cattle host. Bovine TB has been documented across all continents except Antarctica leading to the general assumption that bovine disease is present where cattle are found. Historical data suggest that bovine TB could have emerged in Europe, and thereafter, it would have been distributed to European colonies by exportation of infected animals. Aside from being an economic problem for farmers, bovine TB has historically represented a primary cause of human TB through consumption of unpasteurized milk. Pasteurization of milk and epidemiology control measures that implies test and slaughter of infected cows have served to substantially reduce the incidence of *M. bovis*-related TB in humans.[26] Today, three major *M. bovis* complexes (African 1, African 2, and European 1) exist with characteristic phylogeographical distribution.[27] These complexes are likely the result of founder effect by which a clone is introduced in a new territory and subsequently flourishes in the absence of competition. The African 1 (Af1) subtype is characterized by the loss of RDAf1 and is frequently found in West-Central Africa. However, typing has demonstrated that each country has unique population structures indicative that transhumance movements have not substantially contributed to disperse and homogenize the Af1 population. The African 2 (Af2) subtype is characterized by the loss of RDAf2. Af2 isolates are dominant in East Africa and similar to Af1 strains, there exist unique population structures suggesting that population mixing is infrequent. Altogether Af1 and Af2 are mutually exclusive since they do not share phylogenetic history and rarely found out of Africa. The European 1 (Eu1) subtype lacks a specific chromosomal deletion (RDEu1) and it is worldwide distributed in America (with the exception of Brazil), East Asia, Australia, New Zealand, South Africa, and the British Isles. The dominant presence of Eu1 in the British Isles and British colonies suggests that Eu1 subtype was distributed from the former. This hypothesis is supported by trading records of Hereford beef cattle, which was exported by United Kingdom since the early 19th century. It remains to be answered the *M. bovis* specialization to infect different cattle hosts and the contribution of social, historical, and agricultural factors to the current scenario of bovine TB.[27]

Downloaded From Vet-eBooks.com

2.5 *Other Animal-Adapted Strains:* Mycobacterium caprae *and* Mycobacterium pinnipedii

Disease in goats caused by *M. caprae* is the second major source of animal TB and an important source of zoonosis. Although *M. caprae* was originally considered a subspecies of *M. bovis*, genetic evidences support the elevation of *M. caprae* as a separate member of the MTBC. Indeed, if we look at the deletion-based phylogeny *M. bovis* lacks RD4 region with respect to *M. caprae*, pointing at caprine isolates as ancestors of bovine isolates[28] (Fig. 23.1A). Considering that goat and cattle domestication started 10,000−11,000 years ago, it remains unanswered whether bovine TB is the result of a spillover transmission from infected goats or a scenario where *M. bovis* and *M. caprae* arose independently to infect specific hosts. Genome data from *M. bovis* and *M. caprae* are limited, but the advent of next-generation sequencing is filling the gap in the knowledge of animal-adapted strains. A 2015 study of the genome of three *M. bovis* and one *M. caprae* field isolates has served to validate the ancestry of *M. caprae* relative to *M. bovis* by a genome-based phylogeny.[29]

Mycobacterium pinnipedii infects marine mammals (seals and sea lions).[30] It has been reported that sea mammals could have played a role in transmitting TB to humans in the Americas across the ocean. Phylogenetic data indicate that modern strains of *M. tuberculosis* from America are closely related to those from Europe (L4) supporting the assumption that TB in the New World was introduced during colonization. However, archaeological evidences of TB in pre-Columbian and Peruvian human mummies are incompatible with this notion since these 1000 years old skeletal samples were present before European contact. Genome sequencing revealed that an MTBC member closely related to *M. pinnipedii* was causative of human TB in these ancient Peruvian population (Fig. 23.1D). This scenario is supported by zoonotic transmission of *M. pinnipeddi* to humans and also with the transmission of *M. pinnipedii* to Australian seal colonies through transatlantic movements.[19]

3. Evolution of the *Mycobacterium tuberculosis* Complex From a Genomic Perspective

The growing availability of mycobacterial genomes has been useful to understand which driving forces have shaped the MTBC evolution. The strict clonality and the reductive evolution by accumulating genetic deletions might be indicative of bacterial adaptation to the intracellular environment. As an example, the obligate and noncultivable intracellular pathogen *M. leprae* has a drastic gene reduction resulting in barely 1600 ORF compared to 3900 ORF in *M. tuberculosis*. On the other side, if we hypothesize that the ancestor of MTBC species was an environmental bacterium, HGT is likely to have occurred between this ancestor and other microorganisms living in the same environment. A study based in determination of G + C content, codon usage, and genomic signatures identified genomic islands present in the *M. tuberculosis*

Downloaded From Vet-eBooks.com

genome but absent from closely related species not belonging to the MTBC (*Mycobacterium marinum* and *M. ulcerans*).[31] Importantly, not all *M. canettii* strains carry these genomic islands, suggesting that they could have been acquired by HGT before the speciation of the MTBC. A large 13 kb genomic island coding for 15 ORF encompasses the rv1041c-rv1055 genes, and some genes from this HGT region are involved in virulence either in cultured macrophages or immunodeficient mice, pointing as a pathogenicity island.[31] Fixation of an HGT-transferred region frequently relies upon conferring a selective advantage to the recipient bacteria. HGT transfer in the MTBC ancestor might have helped this bacterium to invade and persist within environmental protozoan phagocytes, which would have evolved to survive within mammalian phagocytic cells. This notion is supported by the observation that pathogenic but not avirulent mycobacteria survive and persist into amebas.[32] Further validating the assumption that the ancestor of MTBC evolved through HGT, it has been demonstrated that different subtypes of *M. canettii* vary in the presence and location of DNA regions, conferring a composite, mosaic-like structure to the ancestor of the tubercle bacillus.[17]

The split between MTBC and its ancestral progenitor is characterized by a genetic bottleneck that reduced to a minimum the genetic diversity. Speciation in the MTBC entails a host preference and the strict clonality in the MTBC could be either an adaptation or a consequence of pathogenicity in a restricted host population. The evolutionary scenario of the MTBC indicates a reductive evolution (Fig. 23.1A and C) suggesting the fixation of different polymorphisms during speciation.[12,15]

However, despite the prevailing view that MTBC genomes have evolved by genetic decay and polymorphism fixation, studies have revealed large-scale duplications in *M. tuberculosis*. It was initially thought that genome duplications were exclusive of the Beijing (L2) lineage, but our cumulative genomic knowledge has allowed the identification of large genome duplications also in L4 strains. The finding that the rv3128c-rv3644 region is prone to duplication events in *M. tuberculosis* might suggest a selective advantage for strains carrying duplications.[33] Supporting this fitness advantage, it has been demonstrated that drug resistance and virulence is associated with gene duplication events in different bacteria.[34] However, it has been also documented the opposite effect for gene duplication; the finding that a 350 kb duplication in L2 strains is selected when bacteria are grown in vitro, but it results in lower virulence for mice argue against a overall fitness advantage for gene duplication events.[35] Gene duplication also occurs in the H37Rv laboratory strain of *M. tuberculosis*, indicative that these polymorphisms are not exclusive of human isolated strains, and consequently, it is an inherent feature of the *M. tuberculosis* chromosome. Although boundaries of some duplicated regions are flanked by repetitive or transposable elements, the precise mechanism for gene duplication remains elusive.

Another mechanism contributing to genome evolution in the MTBC are deletions mediated by the mobile element IS*6110*. This transposon is exclusively present in MTBC members in variable number and localization making it an invaluable tool for molecular epidemiology of TB.[5] The presence of two adjacent IS*6110* copies can lead to recombination with the concomitant loss of the genomic region between them.[36] In order to detect or predict IS*6110*-mediated deletions, it is required to

Downloaded From Vet-eBooks.com

know the precise insertion sites across the chromosome. This task is hardly accomplished by current next-generation sequencing methodologies due to the repetitive nature of IS*6110*, being necessary to locate insertion sites by molecular biology methods. In addition to contribute to the genome plasticity of MTBC members, the mobile element IS*6110* has a function as a mobile promoter, and this effect is more prominent in an intracellular environment.[37] Since transposition frequency in *M. tuberculosis* is higher than the mutation rate in this species, it is expected that IS*6110*-mediated changes (either deletions or increased expression of flanking genes) are important driving forces in the evolution and host adaptation of the MTBC.

4. Evolution in the Laboratory Environment and In Vitro Attenuation of Bacteria From the *Mycobacterium tuberculosis* Complex

In terms of pathogenesis, it is tempting to think that bacterial evolution is exclusively focused on host adaptation. However, in the context of the MTBC, there are two interesting examples of bacterial attenuation in the laboratory that deserve attention: H37 dissociation and development of BCG. These historic landmarks have provided mycobacteriologists with laboratory reference strains and a vaccine against TB.

The *M. tuberculosis* H37 parent strain was originally isolated from a patient with chronic pulmonary disease by Edward R. Baldwin in 1905. In 1934, Steenken and collaborators reported that repeated subculture of the H37 strain in solid egg media at pH = 6.2 resulted in two different colony morphologies. The passages continued until two stable variants were obtained. One of these variants was completely attenuated when inoculated in guinea pigs, a highly sensitive model of TB. Accordingly, these strains were designated H37Rv (v for virulent) and H37Ra (a for avirulent).[38] Both strains were maintained at Trudeau Institute (New York) for many years and were later deposited in the American Type Culture Collection. Historically, H37Rv and its attenuated counterpart H37Ra have been widely used as reference strains for virulence studies of *M. tuberculosis* since 1940. The genetic basis for H37Ra attenuation remained unclear until 2008 when its genome sequence was publicly available 10 years after deciphering the genome sequence of its H37Rv counterpart. When compared with H37Rv, H37Ra acquired multiple point mutations, deletions, and/or genomic rearrangements during in vitro passage. Although the precise role for every polymorphism in H37Ra attenuation is elusive, it has been reported that a point mutation resulting in a single aminoacid substitution (Ser219Leu) in the *phoP* gene have significantly contributed to H37Ra attenuation.[39] As a consequence of this mutation, the PhoP-PhoR virulence system is nonfunctional in H37Ra, and accordingly, this strain lacks immunomodulatory lipids and does not secrete ESAT-6.[40,41] However, even if these phenotypes are key contributors to the avirulence of H37Ra, they fail to explain the complete attenuation of this strain. Thus other polymorphisms arisen during in vitro passages (for example, those leading to loss of the phtiocerol

Downloaded From Vet-eBooks.com

dimycocerosate (PDIM) virulence lipid) likely contribute together with the *phoP* mutation to the avirulent phenotype of H37Ra.

The BCG vaccine receive this name from Bacille de Calmette et Guérin from the scientists who developed this strain almost a century ago. In 1900, Albert Calmette and Camille Guérin at the Pasteur Institute in Lille began their research for a TB vaccine. In an attempt to avoid clumping of mycobacteria in liquid broth, they eventually discovered that addition of ox bile to the medium resulted in lowering the virulence of the culture. They applied this knowledge to attenuate an *M. bovis* strain isolated in 1902 from a cow with TB mastitis. In 1908, Calmette and Guérin started successive cultures of *M. bovis* in bile, glycerin, and potato medium. By 1913, their plan to initiate a vaccination trial in cattle was interrupted by World War I and they continued *M. bovis* subcultivation throughout the German occupation of Lille. In 1919, after 230 passages, they obtained a bacterium that resulted attenuated in rabbits, guinea pigs, cattle, and horses. This paved the way for an human vaccination trial in infants who were victims of TB mainly through the consumption of untreated cow milk. This trial started in Paris in 1921 with a severe disease reduction in vaccinated children compared to the unvaccinated controls.[42] After these encouraging results, BCG lots were distributed in 1924 to different countries where they continued being propagated in the same original way as at the Pasteur Institute. Subcultivation was the unique mean available to maintain the vaccine properties of BCG and to avoid the reversion to a virulent state. In 1960−70s and promoted by the availability of low temperature freezers and/or lyophilizers, master seed lots of BCG were prepared worldwide. Today, we know that during its worldwide subcultivation, BCG continued the in vitro evolution leading to the accumulation of specific deletions and polymorphisms that have ultimately resulted in the presence of BCG substrains with different vaccine potential.[43] Almost a century after its initial construction, we know the attenuation basis of BCG. Compared to *M. bovis*, BCG has lost more than 100 genes most of them grouped in RD regions.[44] Specifically, deletion of the 9.5 kb RD1 region in BCG is thought to have substantially contributed to BCG attenuation (Fig. 23.1A). The RD1 region-encoding Rv3871-Rv3879 includes ESAT-6 and CFP10, known to be key virulence factors of virulent mycobacteria.[45] In addition, next-generation sequencing of 14 BCG strains in conjunction with transcriptional and proteomic analysis have served to reveal the profound differences existing between BCG strains[46] as a consequence of a 100 year in vitro evolution.

Altogether, these examples emphasize the effect of laboratory practices in the evolution of MTBC strains. It is well documented that during extended periods of in vitro culture, *M. tuberculosis* undergoes a spontaneous loss of PDIM known to be a key virulence lipid. Consequently, selection of PDIM-negative colonies might bias experimental findings.[47] In the same line, a 2014 study has identified a polymorphism in the promoter region of the *whiB6* gene that alters the regulation of this gene. Importantly, this polymorphism that results in a decreased secretion of ESAT-6 is exclusive of *M. tuberculosis* H37Rv and H37Ra being absent in the remaining members of the MTBC.[48] Aside from the role of WhiB6 in regulating ESAT-6 secretion, other WhiB6-regulated functions might bias the experimental findings when using a single reference strain, making advisable the use of several strains in laboratory research.

Downloaded From Vet-eBooks.com

5. Short-Term Evolution of *Mycobacterium tuberculosis* During Infection, Drug Treatment, and Disease

The current phylogenetic tree of the MTBC is the result of a long host−pathogen coevolution during thousands of years. This speciation in the MTBC has likely occurred by fixation of polymorphisms conferring selective advantages to infect a specific host. Thus it is tempting to assume that different polymorphisms arise during a single infection−disease−transmission cycle and those allowing a better bacterial survival in the host population will become fixed and maintained in a bacterial lineage. During the course of infection, members of the MTBC face numerous stresses and the ability of the bacillus to overcome these barriers will likely result in the fixation of advantageous alleles. However, identification of these polymorphisms has remained hampered until recently due to two main reasons: the broad consideration that the MTBC is strictly clonal and the technical limitation to sequence different bacterial populations from a single host. Today, the use of sensitive, high-density sequencing techniques has allowed the transition from a scenario where each host is represented by a single MTBC strain isolated in a pure culture to a scenario where each host is known to harbor a diverse bacterial population.

A pioneering study in 2011 using the cynomolgus macaque model aimed to identify polymorphisms in *M. tuberculosis* during active, latent, or reactivated disease by sequencing 33 strains representative of these infection stages. Although no insertions or deletions were detected, 14 SNP were identified and validated by traditional Sanger sequencing. These SNPs were not present in the initial inoculum suggestive that they arose during the course of the infection. The mutational capacity calculated as the mutation rate per generation was comparable in active, latent, and reactivated bacteria. Within a specific lesion, SNPs were independent or shared between the sequenced strains, which indicates that some SNPs accumulate within lesions over the course of the infection. In addition, the study suggests that the polymorphic pattern is the result of oxidative DNA damage during the adaptive immune response to the infection. Specifically, a significant proportion of SNPs corresponded to cytosine deamination (GC > AT) or formation of 8-oxoguanine (GC > TA) consistent with the oxidative environment in the phagolysosome.[49]

In a 2015 study, genome data from *M. tuberculosis* isolated during the course of infection in five different patients were interrogated. The experimental design included *M. tuberculosis* strains from different lineages and different susceptibility profiles to anti-TB drugs and patients from different origins. A series of statistical measures to quantify the mutation rate within a patient revealed that several SNP arose across the course of the infection. Diversity intrapatient varies dramatically, likely as a function of disease severity. When examining the SNP diversity intra and between-patient, it was observed that genes involved in the regulation, synthesis, and transport of immunomodulatory lipids were prone to SNP accumulation.[50] Given the role of these lipids in the interplay with the host immune system, it is not surprising that genes related to their synthesis are linked to a positive selection. Although this study does not represent either the global host population or the broad MTBC phylogeny or the entire bacterial

Downloaded From Vet-eBooks.com

population in the infected lung, it reinforces the notion of intrapatient diversity. Overall, this kind of studies starts to delineate the evolutionary forces that can turn a single bacterium into a widespread lineage, but future works are needed to decipher the biological significance of these polymorphisms arisen in vivo in the context of host–pathogen coevolution.

A particular scenario of MTBC adaptation to the host is related to drug treatment of infected patients. Since anti-TB drugs represent an immediate threat for bacterial survival, polymorphisms conferring drug resistance will be rapidly fixed in the population. The evolutionary forces governing host adaptation and drug resistance operate at different timescales and they affect different alleles. A study collected a total of seven sputum samples from three patients and examined genetic mutations during different stages in the development of drug resistance. Surprisingly, in all seven samples, it was found a high diversity to adapt to antibiotic stress that as many as four to five resistant mutants were detected in a single sputum, suggesting that drug resistance results from a heterogeneous genotype. However, the study suggests that although multiple resistant mutations coexist in the host, ultimately only a single resistance mutation will be fixed in the population.[51]

Another question is why some *M. tuberculosis* strains are preferentially associated with the acquisition of resistance to multiple drugs, which poses a dangerous threat against TB control programmes. Epidemiological data indicate that isolates from the East-Asia lineage (L2), which include Beijing strains, are associated with an increased risk of drug resistance in several countries. A set of experiments demonstrated that L2 strains acquire resistance to drugs in vitro more rapidly than strains from L4, which is the result of the higher basal mutation rate of the L2 lineage. Consequently, it is proposed that L2 strains acquire drug resistance mutations even before the contact with the antibiotic. Extrapolation of this result alongside with estimate of mutation rate in humans suggest that individuals infected with L2 strains are at a notably increased risk of acquiring multiple drug resistance before treatment compared to individuals infected with other strains.[52] Considering the worldwide expansion of Beijing-L2 strains, this finding has extraordinary implications for diagnosis and treatment of these isolates. Since a higher bacterial burden is associated with a higher number of potential drug-resistant bacteria, improvements in diagnostic tools, and treatment regimes can help to limit the emergence of drug resistance.

In the context of the evolution of drug-resistant strains, it remains to be elucidated why these bacteria successfully survive in the host population even if the drug resistance phenotype is frequently associated with a reduced fitness compared to drug-susceptible strains. However, it is observed that some *M. tuberculosis* isolates resistant to the first-line drug rifampicin show no fitness reduction compared to susceptible strains. By comparing the genomes of rifampicin-resistant clones with their corresponding susceptible isolates recovered from the same patient at an earlier time point, it was identified a set of potential compensatory mutations in RNA polymerase genes (the target of rifampicin). This hypothesis was confirmed by genome sequencing of in vitro-evolved strains and the finding that a significant proportion of these putative compensatory mutations lied in the α and β' subunits of the RNA polymerase. The presence of these potential compensatory mutations was associated with an increased

Downloaded From Vet-eBooks.com

fitness in vitro. In addition, upon examination the relative frequency of these compensatory mutations across patient populations, it was also possible to correlate the presence of these mutations with an increased fitness within the host.[53] Further work is needed to determine whether strains carrying compensatory mutations are transmitted and maintained successfully in the population. Additional studies will be necessary to determine the set of compensatory mutations for other anti-TB drugs and the fitness balance of these strains especially in the context of multidrug resistance.

An aspect frequently unnoticed when studying MTBC adaptation to their hosts is referred to transposition of IS*6110* during infection. As described earlier, this transposon has a role not only in shaping the chromosome of MTBC members, but also it functions as a mobile promoter. Although the study of IS*6110* is constricted to molecular epidemiology, very few studies have provided insights into the biological function of IS*6110* transposition in the MTBC and/or the transposition rate within the host. However, upon examination of epidemiologic data, it has been reported that IS*6110* has an unusually high mutation rate (7.9×10^{-5} transposition events per site per generation) compared to the basal nucleotide mutation rate in the MTBC ($\approx 10^{-10}$ to 10^{-9} events per nucleotide per generation). Transposition rate is even higher than the mutation rate in hypermutator strains carrying a defective DNA repair machinery (10^{-7} to 10^{-6} events per nucleotide per generation).[54] These observations strongly suggest that transposition of IS*6110* is at positive selection within patients and highlight the importance of IS*6110* as an epidemiological marker due to its high biological variability. However, provided that IS*6110* produces a replicative transposition resulting in duplication of IS*6110* elements, it has been suggested that accumulation of IS*6110* across the chromosome might have a negative effect either by inactivating essential genes or by mediating deleterious genomic deletions. On this basis, it has been suggested that IS*6110* copy number should be somewhat regulated in order to avoid damage to the bacterial host.[55]

6. Adaptive Cues of the *Mycobacterium tuberculosis* Complex As the Most Successful Pathogens

Today, *M. tuberculosis* is present in one every three humans and *M. bovis* is present everywhere where cattle is present. These data illustrate the view that bacteria from the MTBC are the most successful pathogens probably as a result of a long and intimate host–pathogen coevolution. Further, different *M. tuberculosis* and *M. africanum* lineages exhibit characteristic phylogeographical distributions and this is also applied to *M. bovis* subtypes. One might argue that interspecies polymorphisms likely account for their specific host tropism. In favor of this hypothesis, infection of humans with *M. bovis* rarely result in the transmission of the bovine bacillus and the converse is also true, *M. tuberculosis* produces less severe pathology in cattle than *M. bovis*. A recent study in 2014 has paved the way to assign biological roles for inter species polymorphisms and it opens the door to the possibility that these polymorphisms have shaped the host–pathogen range of the MTBC. This study demonstrated

Downloaded From Vet-eBooks.com

a deleterious effect of a Gly71Ile mutation in PhoR from *M. africanum* L6 and from L8 (animal-adapted) species (Fig. 23.1A). Considering the key role of the *phoP-phoR* two-component system in *M. tuberculosis* virulence through the control of immunomodulatory lipids and the secretion of ESAT-6, it was observed that *M. africanum* L6 and *M. bovis* have a down-regulated PhoP regulon, and consequently, they are unable to produce immunomodulatory lipids. Unexpectedly, these strains secreted ESAT-6 independently of the PhoR mutation, due to a compensatory mechanism in order to maintain the required virulence fitness to infect their respective hosts. The absence of immunomodulatory lipids in *M. bovis* might represent an evolutionary step that reduced the ability of this species to transmit between humans. This notion is supported by the fact that TB disease and transmission between humans is greatly benefitted from the onset of an inflammatory response. In addition, this study uncovered the molecular mechanism by which a rare *M. bovis* isolate ("B strain") is able to transmit between humans due to the insertion of an IS*6110* element upstream the *phoP* gene. Since IS*6110* behave as a mobile promoter, this insertion resulted in an increased expression of *phoP*. Consequently, the *M. bovis* "B strain" carries a deregulated *phoP* gene that results in the production of PhoP-dependent phenotypes despite carrying a nonfunctional PhoR. Production of immunomodulatory lipids in the "B strain" probably contributes to the successful human-to-human spread of this rare *M. bovis* isolate as inferred by the higher virulence of this strain in animal models.[22] Altogether, these findings highlight the importance of *phoP-phoR* polymorphisms in shaping host–pathogen adaptation and encourage mycobacteriologists to profound in the implications of other interspecies polymorphisms.

MTBC bacteria persist and progress in their respective hosts despite development of immune responses. Upon engulfment of MTBC bacteria, phagocytic cells are able to present in cell-surface receptors epitopes derived from proteolysis of MTBC antigens. The subsequent recognition of these epitopes by T $CD4^+$ and $CD8^+$ lymphocytes results in antigen-specific immune responses. This can lead to the erroneous assumption that similarly to other pathogens, MTBC members are subjected to immune escape and antigenic proteins tend to be hypervariable and subjected to diversifying selection. At a first glance, clonality in the MTBC argues against this hypothesis. After exploration of 21 representative MTBC genome sequences, it was demonstrated that the ratio of the rates of nonsynonymous and synonymous substitutions (dN/dS) was significantly lower for essential genes than for nonessential genes. Importantly, dN/dS ratios for T-cells antigens were even lower than for essential genes. Overall, MTBC antigens are under purifying selection as strong as, or perhaps even stronger than that of essential genes.[56] The lack of immune evasion reflects that members of the MTBC have developed different strategies for immune subversion and these pathogens probably benefit from being recognized by the host-immune system. Adaptive immunity mediated by recognition of MTBC epitopes results in containment of the infection during long periods. This latent infection with subsequent reactivation to disease is characteristic of human disease and could represent an evolutive strategy to ensure the pathogen transmission to future generations of susceptible hosts.

In line with this observation, it has been suggested that phylogenetic diversity between MTBC lineages would be reflected in a corresponding diversity when eliciting

Downloaded From Vet-eBooks.com

immune responses. Examination of the macrophage inflammatory response after infection with *M. tuberculosis* or *M. africanum* representative of L1−L6 lineages resulted in lineage-specific profiles. Evolutionary modern lineages (L2, L3, L4) elicited lower inflammatory responses than evolutionary ancient lineages (L1, L5, L6).[57] Since development of disease and the subsequent transmission to new hosts is benefited from inflammatory responses, this result has outstanding implications to understand the host−pathogen evolution of the MTBC. Ancient lineages with higher inflammatory responses are associated with a containment of infection and longer latency period. Conversely, the lower inflammatory phenotype in modern lineages is associated with an early progressive disease. To understand the biological significance of this finding, it is necessary to critically observe which host population is preferentially targeted by each MTBC lineage. Ancient strains are associated with low-density countries and mainly infect rural populations in Africa. These nomadic and low-density settlements resemble a scenario of hunting, gathering, and pastoralism characteristic of Palaeolithic and Neolithic period. By contrast, modern strains are frequently found in crowded countries of Europe, India, China, and America. This lifestyle associated with the human demographic explosion was boosted by the modern Industrial Revolution starting in the 18th century.

It has been proposed an "ecological theory" to explain the evolution of the MTBC. This theory predicts that when virulence is positively correlated with transmission, as is the case in TB, access to a larger number of susceptible hosts favors higher virulence and a shorter latency period. Most of the coevolutionary history of MTBC with humans has occurred when human population densities were low. During this period, infection with ancient strains would have guaranteed a latent infection that after reactivation decades later would enable access to a new birth cohort of susceptible hosts. This way, ancient lineages of the MTBC are not disproportionately virulent in order to avoid decimation of the susceptible population. By contrast, living in crowded countries emerging from the 18th century, would have enabled the accessibility to a large number of susceptible hosts. This lifestyle might have selected more virulent strains without risk to decimate their hosts. We can hypothesize that modern lineages are more evolutionary advantaged than ancient lineages since they are less geographically constrained. Hence, from an ecological point of view, ancient strains might be referred to as "specialists," while modern lineages might represent "generalists."[20] Understanding the evolutive strategies of the MTBC may help to predict future trends in the spread of the disease and consequently to anticipate control measures. As an example, according to the ecological theory, the increasing number of the African population will probably entail a displacement of ancient strains by strains belonging to modern lineages in the future. Indeed, an epidemiologic study in the Gambia showed that strains from modern lineages were three times more likely to cause active disease than members from ancient lineages.[23]

Irrespective of the individual adaptations of different lineages to specific hosts, MTBC members share common virulence mechanisms. It has been extensively documented the role of ESAT-6 and its cognate partner CFP10 in promoting mycobacterial virulence. These proteins are not only conserved between members of the MTBC, but also in other related pathogenic species such as *M. leprae* or *M. marinum* causative of

Downloaded From Vet-eBooks.com

human leprosy and fish-to-human zoonosis respectively. Conversely, the BCG vaccine lacking these proteins or *phoP* mutants unable to secrete ESAT-6 result attenuated. A pioneering study demonstrated that 48 h after macrophage infection, *M. tuberculosis* and *M. leprae* translocated from the harsh phagolysosome to the gentle cytosol. This translocation is impaired in the BCG vaccine, which led to the demonstration that phagosomal escape requires a proper production and secretion of these proteins.[45] Subsequent studies went deeply into the mechanism of this phagolysosomal rupture mediated by ESAT-6/CFP10 secretion and linked cytosolic escape with virulence of *M. tuberculosis* and *M. marinum*.[58,59] A 2013 study has demonstrated that in addition to mediate phagolysosomal escape, the RD1 region (containing ESAT-6 and CFP10) plays a role in inducing apoptosis on infected cells. This apoptotic mechanism promotes cell-to-cell spread of virulent mycobacteria and enables colonization of neighboring cells. Accordingly, the BCG vaccine or new generation vaccines based on *phoP* mutants, unable to produce or secrete ESAT-6, respectively, are inefficient in inducing apoptosis or colonizing new cells.[60] In sum, for MTBC members that does not produce classical toxins, we can consider ESAT-6/CFP10 as the main virulence determinants.

7. Pending Questions and Concluding Remarks

Knowledge about mycobacterial diversity of the MTBC has led to propose an evolutive branding of MTBC lineages. However, we still do not know which factor(s) contribute to the host preference in the MTBC. This question is hardly accomplished since the natural hosts of the MTBC are humans, cows, goats, or seals among others. This implies a handicap in animal experimentation since laboratory models (mainly mice, guinea pigs, and nonhuman primates) does not reflect this host variability. Further, these models attempt to characterize virulence traits rather than measuring the transmissibility of MTBC strains, a key phenotype of the tubercle bacillus to ensure survival in the host population. It is also likely that yet unknown host factors have also contributed to shape the current phylogeographical distribution of the MTBC. In this respect, it has been documented a correlation between the *NRAMP1* human allele (associated with TB resistance) and the duration of urban settlements.[61] Other polymorphisms (*IFNG*, *NOS2A*, *MBL*, *VDR*, and some *TLR*) have been associated with susceptibility to TB in case−control association studies.[62] A future linkage between MTBC genomes and representative human haplogroups might infer host susceptibility factors. In a near future, these strategies will allow personalized treatments or even reconsider vaccination strategies.

The advent of next-generation sequencing techniques has proven useful to delineate inter- and between-species polymorphisms in the MTBC. However, little works have gone deeply on the biological significance of these variations. Nevertheless, 2015 studies have succeeded in reconstructing the phylogenetic evolution of the MTBC based on the biological implications of polymorphisms in the *phoP-phoR* virulence system.[63] Similar works are needed to study the functional consequences of the about

Downloaded From Vet-eBooks.com

2000 polymorphisms existing between MTBC species. We should be aware that inherent difficulties in manipulation of mycobacteria and the slow growth of MTBC strains frequently hamper these genetic approaches.

The debate to estimate the MRCA of the MTBC is a consequence of the weakness of experimental evidence. These estimates frequently assume a constant mutation rate over time and over different lineages. This observation was partly validated using a macaque model showing similar mutation rates during in vitro culture and latent and active TB. However, we cannot rule our differences in mutation rates between different lineages of the MTBC. Indeed, it was latter demonstrated that L2 strains exhibit higher mutation rates than L4 isolates in vitro. It is at present unknown whether this finding is extrapolated to the TB infection. Further, as observed with IS*6110* transposition, it might be also possible that some mutations are under positive selection during TB infection and disease. Future interrogation of these questions together with the application of next-generation sequencing to ancient DNA will surely help to reconstruct the origin of the MTBC.

TB, unlike other infectious diseases, is benefited from efficacious immune system recognition. The hyperconservation of T-cell epitopes likely plays a key role in development of a latent TB infection. This latency and reactivation to active disease of the MTBC resembles the lambda phage cycle by promoting prophage transmission for many bacterial generations until a stress response triggers production of virulent phage particles. However, the environmental signal(s) that mediate TB reactivation are poorly understood. Several studies have deciphered the phagolysosomal signals encountered by *M. tuberculosis* by examining transcriptional profiles during macrophage infection. These studies have inferred that phagolysosomal environment is nitrosative, oxidative, functionally hypoxic, carbohydrate poor, and capable of perturbing the pathogen's cell envelope.[64] Other approaches have addressed the time-dependent transcriptional profiles during intracellular infection[65] and studies reported in 2013 have provided unprecedented level of detail about the transcription networks of *M. tuberculosis*.[66] In the future, combination of genomic, transcriptomic, proteomic, and metabolomic data from animal infection models will hopefully delineate the host–pathogen interplay during both the latency period and reactivation.

References

1. Jacobs Jr WR, Tuckman M, Bloom BR. Introduction of foreign DNA into mycobacteria using a shuttle phasmid. *Nature* 1987;**327**(6122):532–5.
2. Martin C, Timm J, Rauzier J, Gomez-Lus R, Davies J, Gicquel B. Transposition of an antibiotic resistance element in mycobacteria. *Nature* 1990;**345**(6277):739–43.
3. Pascopella L, Collins FM, Martin JM, Lee MH, Hatfull GF, Stover CK, et al. Use of in vivo complementation in *Mycobacterium tuberculosis* to identify a genomic fragment associated with virulence. *Infect Immun* 1994;**62**(4):1313–9.
4. Otal I, Martin C, Vincent-Levy-Frebault V, Thierry D, Gicquel B. Restriction fragment length polymorphism analysis using IS6110 as an epidemiological marker in tuberculosis. *J Clin Microbiol* 1991;**29**(6):1252–4.

Downloaded From Vet-eBooks.com

5. van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* 1993;**31**(2):406−9.
6. Mahairas GG, Sabo PJ, Hickey MJ, Singh DC, Stover CK. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *J Bacteriol* 1996; **178**(5):1274−82.
7. Philipp WJ, Nair S, Guglielmi G, Lagranderie M, Gicquel B, Cole ST. Physical mapping of *Mycobacterium bovis* BCG pasteur reveals differences from the genome map of *Mycobacterium tuberculosis* H37Rv and from *M. bovis*. *Microbiology* 1996;**142**(Pt 11): 3135−45.
8. Philipp WJ, Poulet S, Eiglmeier K, Pascopella L, Balasubramanian V, Heym B, et al. An integrated map of the genome of the tubercle bacillus, *Mycobacterium tuberculosis* H37Rv, and comparison with *Mycobacterium leprae*. *Proc Natl Acad Sci USA* 1996;**93**(7): 3132−7.
9. Brosch R, Gordon SV, Billault A, Garnier T, Eiglmeier K, Soravito C, et al. Use of a *Mycobacterium tuberculosis* H37Rv bacterial artificial chromosome library for genome mapping, sequencing, and comparative genomics. *Infect Immun* 1998;**66**(5): 2221−9.
10. Gordon SV, Brosch R, Billault A, Garnier T, Eiglmeier K, Cole ST. Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. *Mol Microbiol* 1999;**32**(3):643−55.
11. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 1998; **393**(6685):537−44.
12. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, et al. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci USA* 2002;**99**(6):3684−9.
13. Fabre M, Hauck Y, Soler C, Koeck JL, van Ingen J, van Soolingen D, et al. Molecular characteristics of "*Mycobacterium canettii*" the smooth *Mycobacterium tuberculosis* bacilli. *Infect Genet Evol* 2010;**10**(8):1165−73.
14. de Jong BC, Antonio M, Gagneux S. *Mycobacterium africanum*−review of an important cause of human tuberculosis in West Africa. *PLoS Negl Trop Dis* 2010;**4**(9):e744.
15. Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, et al. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat Genet* 2013;**45**(10):1176−82.
16. Merker M, Blin C, Mona S, Duforet-Frebourg N, Lecher S, Willery E, et al. Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nat Genet* 2015;**47**(3):242−9.
17. Gutierrez MC, Brisse S, Brosch R, Fabre M, Omais B, Marmiesse M, et al. Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. *PLoS Pathog* 2005; **1**(1):e5.
18. Supply P, Marceau M, Mangenot S, Roche D, Rouanet C, Khanna V, et al. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of *Mycobacterium tuberculosis*. *Nat Genet* 2013;**45**(2):172−9.
19. Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* 2014;**514**(7523):494−7.
20. Gagneux S. Host-pathogen coevolution in human tuberculosis. *Philos Trans R Soc Lond B Biol Sci* 2012;**367**(1590):850−9.

Downloaded From Vet-eBooks.com

21. Smith NH, Hewinson RG, Kremer K, Brosch R, Gordon SV. Myths and misconceptions: the origin and evolution of *Mycobacterium tuberculosis*. *Nat Rev Microbiol* 2009;**7**(7): 537−44.
22. Gonzalo-Asensio J, Malaga W, Pawlik A, Astarie-Dequeker C, Passemar C, Moreau F, et al. Evolutionary history of tuberculosis shaped by conserved mutations in the PhoPR virulence regulator. *Proc Natl Acad Sci USA* 2014;**111**(31):11491−6.
23. de Jong BC, Hill PC, Aiken A, Awine T, Antonio M, Adetifa IM, et al. Progression to active tuberculosis, but not transmission, varies by *Mycobacterium tuberculosis* lineage in the Gambia. *J Infect Dis* 2008;**198**(7):1037−43.
24. Blouin Y, Cazajous G, Dehan C, Soler C, Vong R, Hassan MO, et al. Progenitor "*Mycobacterium canettii*" clone responsible for lymph node tuberculosis epidemic, Djibouti. *Emerg Infect Dis* 2014;**20**(1):21−8.
25. Boritsch EC, Frigui W, Cascioferro A, Malaga W, Etienne G, Laval F, et al. pks5-recombination-mediated surface remodelling in *Mycobacterium tuberculosis* emergence. *Nat Microbiol* 2016.
26. Smith NH, Gordon SV, de la Rua-Domenech R, Clifton-Hadley RS, Hewinson RG. Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*. *Nat Rev Microbiol* 2006;**4**(9):670−81.
27. Smith NH. The global distribution and phylogeography of *Mycobacterium bovis* clonal complexes. *Infect Genet Evol* 2012;**12**(4):857−65.
28. Aranaz A, Cousins D, Mateos A, Dominguez L. Elevation of *Mycobacterium tuberculosis* subsp. *caprae* Aranaz et al. 1999 to species rank as *Mycobacterium caprae* comb. nov., sp. nov. *Int J Syst Evol Microbiol* 2003;**53**(Pt 6):1785−9.
29. de la Fuente J, Diez-Delgado I, Contreras M, Vicente J, Cabezas-Cruz A, Tobes R, et al. Comparative genomics of field isolates of *Mycobacterium bovis* and *M. caprae* provides evidence for possible correlates with bacterial viability and virulence. *PLoS Negl Trop Dis* 2015;**9**(11):e0004232.
30. Cousins DV, Bastida R, Cataldi A, Quse V, Redrobe S, Dow S, et al. Tuberculosis in seals caused by a novel member of the *Mycobacterium tuberculosis* complex: *Mycobacterium pinnipedii* sp. nov. *Int J Syst Evol Microbiol* 2003;**53**(Pt 5):1305−14.
31. Becq J, Gutierrez MC, Rosas-Magallanes V, Rauzier J, Gicquel B, Neyrolles O, et al. Contribution of horizontally acquired genomic islands to the evolution of the tubercle bacilli. *Mol Biol Evol* 2007;**24**(8):1861−71.
32. Jang J, Becq J, Gicquel B, Deschavanne P, Neyrolles O. Horizontally acquired genomic islands in the tubercle bacilli. *Trends Microbiol* 2008;**16**(7):303−8.
33. Weiner B, Gomez J, Victor TC, Warren RM, Sloutsky A, Plikaytis BB, et al. Independent large scale duplications in multiple *M. tuberculosis* lineages overlapping the same genomic region. *PLoS One* 2012;**7**(2):e26038.
34. Galagan JE. Genomic insights into tuberculosis. *Nat Rev Genet* 2014;**15**(5):307−20.
35. Domenech P, Rog A, Moolji JU, Radomski N, Fallow A, Leon-Solis L, et al. Origins of a 350-kilobase genomic duplication in *Mycobacterium tuberculosis* and its impact on virulence. *Infect Immun* 2014;**82**(7):2902−12.
36. Sampson SL, Warren RM, Richardson M, Victor TC, Jordaan AM, van der Spuy GD, et al. IS6110-mediated deletion polymorphism in the direct repeat region of clinical isolates of *Mycobacterium tuberculosis*. *J Bacteriol* 2003;**185**(9):2856−66.
37. Alonso H, Aguilo JI, Samper S, Caminero JA, Campos-Herrero MI, Gicquel B, et al. Deciphering the role of IS6110 in a highly transmissible *Mycobacterium tuberculosis* Beijing strain, GC1237. *Tuberculosis (Edinb)* 2011;**91**(2):117−26.

Downloaded From Vet-eBooks.com

38. Steenken W, Oatway WH, Petroff SA. Biological studies of the tubercle Bacillus: III. Dissociation and pathogenicity of the R and S variants of the human tubercle Bacillus (H(37)). *J Exp Med* 1934;**60**(4):515−40.
39. Lee JS, Krause R, Schreiber J, Mollenkopf HJ, Kowall J, Stein R, et al. Mutation in the transcriptional regulator PhoP contributes to avirulence of *Mycobacterium tuberculosis* H37Ra strain. *Cell Host Microbe* 2008;**3**(2):97−103.
40. Chesne-Seck ML, Barilone N, Boudou F, Asensio JG, Kolattukudy PE, Martin C, et al. A point mutation in the two-component regulator PhoP-PhoR accounts for the absence of polyketide-derived acyltrehaloses but not that of phthiocerol dimycocerosates in *Mycobacterium tuberculosis* H37Ra. *J Bacteriol* 2007;**190**(4):1329−34.
41. Frigui W, Bottai D, Majlessi L, Monot M, Josselin E, Brodin P, et al. Control of *M. tuberculosis* ESAT-6 secretion and specific T cell recognition by PhoP. *PLoS Pathog* 2008;**4**(2):e33.
42. Calmette A. Preventive vaccination against tuberculosis with BCG. *Proc R Soc Med* 1931; **24**(11):1481−90.
43. Behr MA. BCG−different strains, different vaccines? *Lancet Infect Dis* 2002;**2**(2):86−92.
44. Brosch R, Gordon SV, Garnier T, Eiglmeier K, Frigui W, Valenti P, et al. Genome plasticity of BCG and impact on vaccine efficacy. *Proc Natl Acad Sci USA* 2007;**104**(13):5596−601.
45. van der Wel N, Hava D, Houben D, Fluitsma D, van Zon M, Pierson J, et al. *M. tuberculosis* and *M. leprae* translocate from the phagolysosome to the cytosol in myeloid cells. *Cell* 2007;**129**(7):1287−98.
46. Abdallah AM, Hill-Cawthorne GA, Otto TD, Coll F, Guerra-Assuncao JA, Gao G, et al. Genomic expression catalogue of a global collection of BCG vaccine strains show evidence for highly diverged metabolic and cell-wall adaptations. *Sci Rep* 2015;**5**:15443.
47. Domenech P, Reed MB. Rapid and spontaneous loss of phthiocerol dimycocerosate (PDIM) from *Mycobacterium tuberculosis* grown in vitro: implications for virulence studies. *Microbiology* 2009;**155**(Pt 11):3532−43.
48. Solans L, Aguilo N, Samper S, Pawlik A, Frigui W, Martin C, et al. A specific polymorphism in *Mycobacterium tuberculosis* H37Rv causes differential ESAT-6 expression and identifies WhiB6 as a novel ESX-1 component. *Infect Immun* 2014;**82**(8):3446−56.
49. Ford CB, Lin PL, Chase MR, Shah RR, Iartchouk O, Galagan J, et al. Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. *Nat Genet* 2011;**43**(5):482−6.
50. O'Neill MB, Mortimer TD, Pepperell CS. Diversity of *Mycobacterium tuberculosis* across evolutionary scales. *PLoS Pathog* 2015;**11**(11):e1005257.
51. Sun G, Luo T, Yang C, Dong X, Li J, Zhu Y, et al. Dynamic population changes in *Mycobacterium tuberculosis* during acquisition and fixation of drug resistance in patients. *J Infect Dis* 2012;**206**(11):1724−33.
52. Ford CB, Shah RR, Maeda MK, Gagneux S, Murray MB, Cohen T, et al. *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nat Genet* 2013;**45**(7):784−90.
53. Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, et al. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat Genet* 2012;**44**(1):106−10.
54. Tanaka MM. Evidence for positive selection on *Mycobacterium tuberculosis* within patients. *BMC Evol Biol* 2004;**4**:31.
55. Tanaka MM, Rosenberg NA, Small PM. The control of copy number of IS6110 in *Mycobacterium tuberculosis*. *Mol Biol Evol* 2004;**21**(12):2195−201.

Downloaded From Vet-eBooks.com

56. Comas I, Chakravartti J, Small PM, Galagan J, Niemann S, Kremer K, et al. Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat Genet* 2010;**42**(6):498−503.
57. Portevin D, Gagneux S, Comas I, Young D. Human macrophage responses to clinical isolates from the *Mycobacterium tuberculosis* complex discriminate between ancient and modern lineages. *PLoS Pathog* 2011;**7**(3):e1001307.
58. Simeone R, Bobard A, Lippmann J, Bitter W, Majlessi L, Brosch R, et al. Phagosomal rupture by *Mycobacterium tuberculosis* results in toxicity and host cell death. *PLoS Pathog* 2012;**8**(2):e1002507.
59. Houben D, Demangel C, van Ingen J, Perez J, Baldeon L, Abdallah AM, et al. ESX-1-mediated translocation to the cytosol controls virulence of mycobacteria. *Cell Microbiol* 2012;**14**(8):1287−98.
60. Aguilo JI, Alonso H, Uranga S, Marinova D, Arbues A, de Martino A, et al. ESX-1-induced apoptosis is involved in cell-to-cell spread of *Mycobacterium tuberculosis*. *Cell Microbiol* 2013;**15**(12):1994−2005.
61. Barnes I, Duda A, Pybus OG, Thomas MG. Ancient urbanization predicts genetic resistance to tuberculosis. *Evolution* 2011;**65**(3):842−8.
62. Moller M, de Wit E, Hoal EG. Past, present and future directions in human genetic susceptibility to tuberculosis. *FEMS Immunol Med Microbiol* 2010;**58**(1):3−26.
63. Broset E, Martin C, Gonzalo-Asensio J. Evolutionary landscape of the *Mycobacterium tuberculosis* complex from the viewpoint of PhoPR: implications for virulence regulation and application to vaccine development. *MBio* 2015;**6**(5).
64. Schnappinger D, Ehrt S, Voskuil MI, Liu Y, Mangan JA, Monahan IM, et al. Transcriptional adaptation of *Mycobacterium tuberculosis* within macrophages: insights into the phagosomal environment. *J Exp Med* 2003;**198**(5):693−704.
65. Rohde KH, Veiga DF, Caldwell S, Balazsi G, Russell DG. Linking the transcriptional profiles and the physiological states of *Mycobacterium tuberculosis* during an extended intracellular infection. *PLoS Pathog* 2012;**8**(6):e1002769.
66. Galagan JE, Minch K, Peterson M, Lyubetskaya A, Azizi E, Sweet L, et al. The *Mycobacterium tuberculosis* regulatory network and hypoxia. *Nature* 2013;**499**(7457):178−83.

Downloaded From Vet-eBooks.com

The Evolution and Dynamics of Methicillin-Resistant *Staphylococcus aureus*

24

M.M.H. Abdelbary[1], P. Basset[1], D.S. Blanc[1], E.J. Feil[2]
[1]Centre Hospitalier Universitaire Vaudois and University Hospital of Lausanne, Lausanne, Switzerland; [2]University of Bath, Bath, United Kingdom

1. Introduction

Staphylococcus aureus is a Gram-positive bacterium that typically resides asymptomatically in the anterior nares and the skin of mammals. Since its discovery in the 1880s, it has been recognized as a major opportunistic pathogen in humans, responsible for various diseases, ranging from minor skin infections to severe bacteremia and necrotizing pneumonia. Before the era of antibiotics, the mortality rate of patients infected with *S. aureus* exceeded 80%.[1] The introduction of penicillin in the early 1940s saved the lives of tens of thousands of wounded allied troops in the Second World War and dramatically improved the prognosis of patients with staphylococcal infections. However, as early as 1942, penicillin-resistant staphylococci were recognized, and these strains arose via the acquisition of a plasmid carrying a gene encoding a penicillinase (β-lactamase). Although the spread of penicillin-resistant *S. aureus* was initially confined to hospital settings, this was quickly followed by the wider dissemination of resistance in the community. By the late 1960s, more than 80% of both community- and hospital-associated *S. aureus* isolates were resistant to penicillin.[2] This pattern is being repeated for methicillin, an alternative semisynthetic β-lactam antibiotic that was designed to resist β-lactamase. Since the introduction of this antibiotic in the 1960s, various hospital-associated methicillin-resistant *S. aureus* (HA-MRSA) clones disseminated worldwide, and virulent community-associated MRSA (CA-MRSA) and livestock associated-MRSA (LA-MRSA) have continued to emerge and spread from the mid-1990s onward.

2. The Staphylococcal Cassette Chromosome *mec*

S. aureus is naturally susceptible to most antibiotics, and resistance is often acquired by the horizontal transfer of genes from intrinsically resistant coagulase-negative staphylococci. These genes are generally located on mobile genetic elements (MGEs), such as plasmids or cassettes.

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00024-X
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

The resistance to methicillin and all other β-lactam antibiotics is conferred by the acquisition of the methicillin resistance gene *mecA*.[3] This gene is carried on an MGE called the staphylococcal chromosome cassette *mec* (SCC*mec*).[4] This MGE is likely to have been introduced into the *S. aureus* population on multiple occasions from related staphylococcal species.[5,6] Several structural variants of SCC*mec* have been described, which differ in their gene content and size (21−67 kb), but share four characteristics. First, they carry the *mec* gene complex (*mec*) that is made up of the methicillin resistance determinant *mecA*, its expression regulatory genes (*mecR1* [promoter] and *mecI* [repressor]), and the insertion sequence(s). Second, they carry the cassette chromosome recombinase gene complex (*ccr*), which consists of genes that are responsible for the mobility of the element. Third, they have characteristic repeated sequences at both ends. Fourth, they integrate into the *S. aureus* chromosome at a site-specific location (*attBscc*), located within *orfX* near the origin of replication.[7−10] Despite these common characteristics, the detailed structure of SCC*mec* elements is highly divergent. In particular, several allotypic differences have been identified in *ccr* and *mec* complexes,[11] as described in the following paragraphs.

***ccr* gene complex**. So far, three distinct *ccr* genes have been described (*ccrA*, *ccrB*, and *ccrC*) in *S. aureus*. While *ccrC* is usually found alone, *ccrA* and *ccrB* are generally found adjacently on the same element. In addition, several allotypes of *ccrA* and *ccrB* have been identified. The presence of these genes and allotypes has been used to distinguish among the eight different *ccr* types that are currently observed (Table 24.1).

***mec* gene complex**. The region of the *mec* gene complex differs among SCC*mec* elements in its composition of regulatory genes (*mecI* and *mecR1*) and/or insertion sequences (IS431 and IS1272). So far, six classes of *mec* gene complexes have been described (A, B, C1, C2, D, and E) in *S. aureus* (Table 24.1).

These differences of *ccr* and *mec* gene complexes have been used to classify the SCC*mec* elements into different types by combining the class of the *mec* gene complex with the *ccr* allotype. To date, 11 major types of SCC*mec* elements (I−XI) have been reported in MRSA strains (http://www.sccmec.org; http://www.staphylococcus.net; Table 24.1). In addition, the major elements have been further classified into subtypes by differences in three regions other than *ccr* and *mec*, which are designated junction or junkyard (J) regions. It is likely that many other variants of SCC*mec* elements will be discovered with increasing typing, especially of isolates from poorly sampled geographic regions (e.g., Africa).[12] Furthermore, coagulase-negative staphylococci (e.g., *Staphylococcus haemolyticus*, *Staphylococcus epidermidis*, and *Staphylococcus hominis*) contain a high diversity of SCC*mec* elements, which might serve as a potential reservoir for *S. aureus*.[6,13,14]

The typing of *SCCmec* elements has become essential for several reasons. First, in combination with the genotype of the *S. aureus* chromosome, the SCC*mec* type is an important characteristic for defining MRSA clones in epidemiological studies and to understand the evolution of these clones.[15] Second, the various SCC*mec* elements also differ in their patterns of antibiotic susceptibility, which have important clinical implications. For instance, SCC*mec* type I as well as type IV−VIII cause only resistance to β-lactam antibiotics. In contrast, the largest SCC*mec* types II and III cause

Downloaded From Vet-eBooks.com

Table 24.1 Major SCC*mec* Elements Identified in *Staphylococcus aureus* From Ref. 11

ccr Gene Complex		*mec* Gene Complex		SCC*mec*
***ccr* Genes**	***ccr* Type**	***mec* Genes**	***mec* Class**	**Type**
ccrA1 and *ccrB1*	1	IS*1272*-*ΔmecR1*-*mecA*-IS*431*	B	I
ccrA2 and *ccrB2*	2	*mecI*-*mecR1*-*mecA*-IS*431*	A	II
ccrA3 and *ccrB3*	3	*mecI*-*mecR1*-*mecA*-IS*431*	A	III
ccrA2 and *ccrB2*	2	IS*1272*-*ΔmecR1*-*mecA*-IS*431*	B	IV
ccrC	5	IS*431*-*ΔmecR1*-*mecA*-IS*431*	C1[a]	V
ccrA4 and *ccrB4*	4	IS*1272*-*ΔmecR1*-*mecA*-IS*431*	B	VI
ccrC	5	IS*431*-*ΔmecR1*-*mecA*-IS*431*	C2[a]	VII
ccrA4 and *ccrB4*	4	*mecI*-*mecR1*-*mecA*-IS*431*	A	VIII
ccrA1 and *ccrB1*	1	IS*431*-*ΔmecR1*-*mecA*-IS*431*	C2[a]	IX
ccrA1 and *ccrB6*	7	IS*431*-*ΔmecR1*-*mecA*-IS*431*	C1[a]	X
ccrA1 and *ccrB3*	8	blaZ-*mec*A*LGA251*-*mec*R1*LGA251*-*mec*I*LGA251*	E	XI

[a]*mec* Class C1 and C2 differ in the orientation of IS431 upstream of *mecA*.

resistance to multiple classes of antibiotics due to the integration of plasmids or transposons carrying multiple resistance genes within these elements.

Several SCC*mec*-typing methods have been developed, among which the most widely used are based on multiplex PCR assays that identify the different *ccr* types and *mec* classes.[15–19] These have a limited number of targets, which may restrict their resolution but can be combined according to the level of discrimination required by the study. Two additional sequence–based typing methods based on the *ccr* gene complex have also been proposed,[17,20] and these are likely to provide further useful data. Although SCC*mec* typing is essential for the characterization of MRSA clones in epidemiological studies, it is only recently that a rationalized nomenclature for the SCC*mec* has been proposed.[7,11]

***mecA* gene homologue (*mecC*).** Discovery of a novel *mecA* gene homologue, called *mecC*, was reported in 2011 in the genome of *S. aureus* strain LGA251 that was isolated from bovine mastitis.[21] MRSA strains harboring *mecC* have subsequently been reported from several European countries, and are associated with multiple host species including humans.[22–26] Similar to the *mecA* gene, *mecC* is located within the SCC*mec* element (SCC*mec* type XI) and inserted into the 3′ region of *orfX*. In addition, several *S. aureus* virulence factors, such as adhesions, and toxins were detected among *mecC* MRSA strains.[27,28] The *mecC* gene has been detected in several staphylococcal

Downloaded From Vet-eBooks.com

and other related bacterial species, although the origin of *mecC* in *S. aureus* remains unclear.

Currently, a broad range of commercial and PCR-based approaches are available for the detection of *mecC* MRSA strains, and these have significant diagnostic value for both human and veterinary public health.[29–32]

3. Evolution of *Staphylococcus aureus* and MRSA

Most detailed studies on the population genetics of *S. aureus* have been performed using MLST (Box 24.1). Based on MLST data, the population of *S. aureus* was classified into related groups of strains defined as clonal complexes (CCs) and isolated sequence types (STs).[33] These CCs are considered as different genetic lineages within the *S. aureus* population and only few differences are detected within groups although the characteristics of MGEs (e.g., SCC*mec*) may vary substantially.[34]

Box 24.1 Common Typing Methods for *S. aureus*

The epidemiology of *S. aureus* has been analyzed by an array of genotypic and phenotypic typing methods. Here, we review the methods that are currently the most widely used:

Pulsed-field gel electrophoresis: Pulsed-field gel electrophoresis (PFGE) is considered as the gold standard for *S. aureus* typing because it shows the highest discriminatory power. This method is based on the restriction of whole DNA with an enzyme that cuts only rarely. The enzyme *SmaI* is generally used for *S. aureus*. Digestion with this enzyme gives between 20 and 50 large fragments (between 10 and 700 kb) that can only be separated using a pulsed gel electrophoresis. Although this method is reproducible within a laboratory, the data can be ambiguous[124] and interlaboratory studies have highlighted the problem of standardization.[125] PFGE standardization can only be obtained with a strict control of all parameters. For example, standardized protocols have been developed for PFGE typing by the American and Canadian CDCs to build nationwide databases.[126,127]

Multilocus sequence typing: Multilocus sequence typing (MLST) is a typing method that combines the sequence of several housekeeping genes, and is essentially a sequence-based version of multilocus enzyme electrophoresis (MLEE).[128] MLST has been designed to analyze and compare genetic variation in worldwide collections of bacterial pathogens. It gives important information about the nucleotide divergence of the core genome, the clonal origin of one group of strains, the recombination rate, and the phylogenetic relationship among strains. The main advantage of this method is that it gives unambiguous data that are reproducible among laboratories. Its limitations are its cost and its relatively

Downloaded From Vet-eBooks.com

Box 24.1 Common Typing Methods for *S. aureus*—cont'd

low discriminatory power that prevent its use for local epidemiology. For *S. aureus*, the amplification and the sequencing of 450−500 bp of the seven genes *arcC*, *aroE*, *glpF*, *gmk*, *pta*, *tpi*, and *yqiL* have been retained.[60] Alleles at each locus are assigned according to differences in nucleotide sequences. The allelic profile of the seven loci defines the sequence type (ST). For example, isolates with the profile 2-3-1-1-4-4-3 belong to ST 239, of which the Brazilian clone is an example. An international database containing more than 3000 isolates and 1600 STs is available at http://www.mlst.net.

spa-typing: *spa*-typing is based on polymorphism of the *spa* locus of *S. aureus*, which codes for the protein A. This locus is highly polymorphic due to an internal variable region of short tandem repeats. It varies not only in numbers but also because of nucleotide substitutions within individual repeats. A *spa* profile is identified by a succession of number representing each individual repetition of the X region. An international database has been created to standardize the nomenclature of the *spa* types (http://spaserver.ridom.de). Several studies have shown the value of this method for *S. aureus* typing.[129−131] However, this method might reflect homoplasy,[34] its discriminatory power is below PFGE,[132−134] and the analysis of *spa* data is not simple.

Double locus sequence typing: We developed a new typing method called double locus sequence typing (DLST) based on the analysis of partial sequences (ca. 500 bp) of the highly variable *clfB* and *spa* genes.[133] This method was shown to be far more discriminatory than *spa*-typing and matched the high resolution of PFGE. In addition, the combination of high typeability and reproducibility with low cost, ease of use, and unambiguous definition of types makes this method promising for epidemiological analyses. It is important to note that although *spa*-typing and DLST investigate polymorphisms in the *spa* gene, these methods do not analyze the same regions of the gene. Therefore the *spa* alleles determined by these two methods are not identical.

Whole-genome sequencing (WGS): Recently, high-throughput or whole-genome sequencing technologies have provided a significantly improved discriminatory power to study the complete genomes of various bacterial pathogens. WGS techniques generate from bacterial samples multiple short reads that can be assembled based on overlapping regions (de novo assembly), and/or mapped to a previously published reference genomes, which then enable the comparison between bacterial strains that genetically diverge at a single nucleotide. Such precise identification and classification of bacterial strains, as well as the parallel sequencing of different bacterial strains in single runs at low costs with a quicker turnaround times have made WGS the most convenient tool for clinical diagnostic investigations in real time and for tracking disease outbreaks.

Downloaded From Vet-eBooks.com

Extensive typing showed that the *S. aureus* population associated with humans consists of 10 major lineages (i.e., CC1, CC5, CC8, CC12, CC15, CC22, CC25, CC30, CC45, and CC51), as well as several other minor lineages (Fig. 24.1).[18,33,35] These lineages have not only been identified using MLST but also using other categories of genes[36] confirming the biological reality of the CCs. These CCs generally have a radial genetic structure with a founder ST surrounded by numerous single locus variants of the founder. This observation highlights that with the exception of MGEs the genetic diversity within each lineage is remarkably low.[34,37] For example, the nonmobile genome of two strains belonging to CC1 (MW2 and MSSA476) differ at only 285 single nucleotide polymorphisms (SNPs) despite one was a PVL-positive MRSA isolated



Figure 24.1 Neighbor-Joining tree of the concatenated sequence of all the STs available in the *S. aureus* MLST database (http://saureus.mlst.net/). The position of the sequence types belonging to major *S. aureus* clonal complexes is indicated by a *colored dot*.

Downloaded From Vet-eBooks.com

in the United States and the other a PVL-negative MSSA isolated in the United Kingdom.[38] The low variability observed within CCs might be explained by recent expansion and/or strong purifying selection. Although the relative contribution of each of these factors is difficult to disentangle, purifying selection was described for several categories of genes, such as the seven housekeeping genes used for MLST typing, core and accessory adhesion genes,[39] as well as many others.[40–42] This suggests that purifying selection is an important factor that acts on the chromosome of *S. aureus* and it probably affects the diversity observed within CCs.

Another important factor is the low rate of homologous recombination within the core genome of *S. aureus*. Although several chromosomal replacement events were described for *S. aureus*,[43,44] this species has been shown to be highly clonal using a variety of genes: MLST,[35] cell surface *sas* genes,[36] cell surface core and accessory (i.e., not present in all the strains) adhesion genes,[39] accessory exotoxin-like genes.[45] For example, using MLST data, it was shown that genetic differences between a single locus variant and its ancestral strains were created 15 times more frequently by a point mutation than by a recombination event.[35] This low rate of recombination can help to explain why the clonal complexes have remained discrete and coherent in the *S. aureus* population, and why the same basic groups tend to be defined regardless of the genes used for typing (with the notable exception of *agr*[46]).

The genetic diversity of MRSA is also known to be much smaller than MSSA[47] (and the most common MRSA isolates belong to only six CCs (i.e., CC1, CC5, CC8, CC22, CC30, and CC45)). In contrast to MSSA, the genetic diversity of MRSA differs considerably among countries and dominant MRSA lineages form distinctive geographic clusters, at least in Europe.[47] This largely reflects the recent origin of many MRSA clones, that is, since the first administration of methicillin in 1961. This means that there has been insufficient time for the MRSA clones to fully homogenize geographically.

3.1 Mobile Genetic Elements

Bacterial genomes can be viewed as two compartments of genes, one comprising "core" genes that are ubiquitously present in all clones of a given species, and the other comprising "accessory" or "noncore" genes that are not present in all isolates in the population, and that have a propensity for horizontal transfer.[48] Whole-genome sequence and microarray data have revealed that about 75% of a typical *S. aureus* genome is present in more than 95% of the strains (i.e., core genome).[49–51] As expected, the majority of genes comprising the core genome are composed of species-specific genes and genes associated with central metabolism and other housekeeping function. In contrast, the gene content of the remaining 25% of the genome varies significantly among strains (i.e., accessory genome).

The accessory genome mostly consists of MGEs, such as bacteriophages, pathogenicity islands, genomic islands, staphylococcal chromosome cassettes (SCCs), plasmids, or transposons. Many of these genetic elements carry virulence genes (e.g., *tst* and *PVL*, which are carried on bacteriophages)[51,52] and are resistant to antibiotics (e.g., *mecA carried in SCCmec*).[33] The gene content of a particular *S. aureus* strain

Downloaded From Vet-eBooks.com

is thus a combination of (1) vertical inheritance of its core genome and (2) horizontal transfer of MGEs, allowing rapid adaptation by loss or gain of virulence and/or resistance genes.[51] Thus, there is a considerable proportion of the genome that is not essential for survival and that contributes to genetic differences between strains. The distribution and horizontal spread of these elements can have important clinical implications and the characterization of these elements is providing insights into how *S. aureus* is evolving and how it causes diseases.

Whole-genome comparisons indicated important variation in the distribution of genomic islands. This suggests that MGEs are readily exchanged in the *S. aureus* population. For example, genome comparison of one MRSA strain with one MSSA strain showed at least five different acquisition/loss events involved in differences in virulence factors and drug resistance.[51] Horizontal transfer of MGEs is also suggested by the phylogenetic distribution of these elements, which does not correlate with the genetic relatedness inferred by MLST. This lack of correlation suggests that mobile elements facilitate the exchange of virulence and antibiotic resistance determinants between *S. aureus* lineages and may lead to rapid changes in the pathogenic potential or drug resistance of strains. In contrast, the sequence of the core genome is remarkably constant.[10,53–56]

Several studies suggested that some toxin genes (e.g., toxic shock syndrome toxin 1 (*tst*), leukocidin DE (*lukDE*), and superantigens (*sea*, *seg*, and *sei*)) are associated with particular lineages (MLST CCs),[57,58] and there is evidence of frequent acquisition and loss of particular elements that is restricted to particular CCs. In 2009, variability of accessory genes, such as resistance, toxin, or virulence genes, was described within two STs (ST5 and ST228) of CC5.[59] However, the biological significance and modalities of this intrastrain variability still need to be clarified.

The importance of understanding the patterns of evolution of MGEs is illustrated by the evolution of the SCC*mec* element. The evolution from MSSA to MRSA involves the acquisition of a SCC*mec* element by an MSSA strain. The exact mechanisms explaining how the SCC*mec* elements enter the *S. aureus* cell are not clearly known. However, the transduction by phages is often postulated.[8] The frequency of transfer of SCC*mec* elements as well as their geographic history is also poorly known. Identical clones have been sampled in different countries suggesting a single *SCCmec* acquisition, followed by clonal spread. Yet, the presence of multiple SCC*mec* types in MRSA suggests multiple introductions into *S. aureus*. Moreover, the occurrence of isolates with identical ST but with different *SCCmec* types indicates that horizontal transfer of SCC*mec* elements is relatively frequent within *S. aureus*.[33] Using MLST, it has been shown that the SCC*mec* element must have been acquired on multiple occasions (at least 20 times) during *S. aureus* evolution.[20,33,36] A previous study based on SNPs discovery on a worldwide collection of ST 5 showed a close association between phylogenetic lineages and geography.[34] These data suggest that geographic spread of MRSA over long distance is a rare event compared with the frequency with which the *SCCmec* is imported locally. Moreover, MSSA strains genetically identical to the predominant MRSA strains have been observed at a local level,[60,61] confirming the possibility of local acquisitions of the SCC*mec*.

Downloaded From Vet-eBooks.com

4. Molecular Epidemiology of MRSA

Epidemiological surveillance of MRSA has been greatly facilitated by the development of molecular typing procedures. The grouping of isolates into clones depends on the typing method used (e.g., PFGE, MLST, *spa* typing, and subsequently, whole-genome sequencing; Box 24.1). MLST provides a robust typing system by grouping related *S. aureus* strains into distinct sequence types (STs) based on the sequences of internal fragments of seven housekeeping genes.[33] However, as MLST is only based on the variation within a very small proportion of the genome, much of the fine molecular microevolutionary details, such single nucleotide polymorphisms (SNPs), genome rearrangements, and small INDELS, remain undetected. Even before the advent of whole-genome sequencing, it was clear that single MLST genotypes often encompassed multiple types as defined by other techniques. For example, ST 239 includes EMRSA 1, 4, 11 and the Brazilian, Portuguese, Viennese, and Hungarian clones, and ST 5 includes the New York/Japan and the Pediatric clones, as defined by PFGE, *spa*-, and/or SCC*mec*-typing.[62] Similarly, two Swiss clones (clone D and G) were indistinguishable by MLST, exhibited identical STs (ST228), SCC*mec* type I and virulence gene content as determined by PCR, yet differed by 16 bands by PFGE.[63] These differences were likely the result of the gain or loss of mobile genetic elements (MGEs), such as phage, which would not be detected by other approaches. In contrast, in other cases, microvariation is detected by MLST and other methods (such as individual SNPs) but not by PFGE. None of the traditional typing methods provided optimal resolution in all cases. The advent of whole-genome sequencing has provided such a "one size fits all" approach, in that it provides unprecedented discriminatory power for epidemiological surveillance, outbreak investigations, and better understanding of the evolutionary dynamics of both the core and noncore genome of MRSA.

Many studies have demonstrated that high frequencies of MRSA within a given location tend to reflect the clonal spread of only one or two clones (e.g., [62–70]). The domain of dominance of specific clones can range in size from a single hospital, single country, or even neighboring countries.[71–73] Analysis of more than 3000 isolates from southern Europe, United States, and South America showed that nearly 70% of them belong to five major pandemic clones, namely the Iberian (ST247-SCC*mec* I), Brazilian (ST239-SCC*mec* III), Hungarian (ST239-SCC*mec* III), New York/Japan (ST5-SCC*mec* II), and Pediatric (ST5-SCC*mec* IV) clones.[18,62,74] The addition of three more clones would essentially encompass northern Europe: the EMRSA-15 (ST22-SCC*mec* IV), EMRSA-16 (ST36-SCC*mec* II), and Berlin (ST45-SCC*mec* IV) clones.[33] Therefore, it was hypothesized that these clones are particularly transmissible and/or well adapted to the hospital environment.[75,76]

The epidemiology of MRSA is highly dynamic, and clonal replacement of predominant clones within a given locale has been widely documented. While cross-sectional studies showed the predominance of one or two clones in a defined setting in the 90s, several longitudinal studies showed the replacement of the predominant clones by others within a decade.[62,77] A very early example was the replacement in England of EMRSA-1 (ST239) by EMRSA-15 and -16.[76] Other ST239 variants (e.g., in

Downloaded From Vet-eBooks.com

particular, the Brazilian clone and the Hungarian clone) have subsequently become very widespread throughout South America, Eastern Europe, and mainland Asia (including both China and the middle East), where this genotype may account for at least 90% of all cases of HA-MRSA. In addition, another pandemic clone replaced the Iberian clone on at least two occasions. It was first replaced by EMRSA-16 in one Spanish hospital while the rate of MRSA among *S. aureus* remained constant,[77] and by the Brazilian clone in one Portuguese hospital.[78] The fact that on both occasions the Iberian clone was replaced might suggest that it lost its epidemic potential during the last decade. Other examples are the complete replacement in a 2-year period of a local clone (ST5-SCC*mec* IV) by the New York/Japan clone (ST5-SCC*mec* II) in a Mexico City hospital (Velazquez-Meza et al., 2004) and the replacement of the Berlin clone by a variant from the New York/Japan clone (ST105-SCC*mec* II) and by the South Germany clone (ST228- SCC*mec* I) in an area of low MRSA incidence in western Switzerland.[63] Although the reasons why some clones replace others are typically unclear, the emergence and replacement of clones might have significant public health consequences as different clones possess differing resistance and virulence attributes.[77–83] For instance, during the 1990s in France, the replacement of the Iberian clone (ST 247-SCC*mec* I) by the Lyon clone (ST8-SCC*mec* IV) resulted in a change of the susceptibility profile to antibiotics, the Iberian clone being less susceptible than the Lyon clone (e.g., to gentamicin and co-trimoxasole).[79]

For regions outside of Europe, North America, and Australia the picture may be different. For example, ST239 that probably emerged in the mid-1960s[37] is a probable major cause of HA-MRSA infection throughout mainland Asia and South America, a geographic region that holds more than 50% of the world's human population.[84] This sequence type always exhibits a variant of the large SCC*mec* type III; however, four cases of HA-MSSA ST239 were detected in China.[85] ST239 is rarely found outside of the hospital setting, which makes its rapid global dissemination, which must have occurred largely through very short transmission chains between hospitals, even more remarkable.

Perhaps of greatest concern is the emergence of specific MRSA clones within the community. Up until the 1990s, MRSA was found to be restricted to hospitals, but the 2000s have witness a dramatic increase in virulent MRSA clones in the community (CA-MRSA).[86] These clones are generally characterized by the presence of a SCC*mec* type IV or V and the phage-borne genes encoding the Panton-Valentine leukocidin (PVL) toxin. This toxin is widely considered to be an important virulence factor, particularly for pediatric infection. Molecular typing has revealed that CA-MRSA clones are distinct from those noted in hospital settings.[70,87] ST80-SCC*mec* IV provides a notable example of an emerging CA-MRSA clone, which is currently restricted to several European communities with low social status (e.g., homeless people). Although the widespread HA-MRSA does not appear to have adapted to the community, it seems that clones that emerge in the community may be able to spread in hospitals. For example, ST8-SCC*mec* IV (generally called the USA300 clone) spread mainly in the United States, initially in the community, but is currently also causing a major burden in hospital settings.[88,89] In countries with low incidence of hospital

Downloaded From Vet-eBooks.com

MRSA, such as northern European countries, CA-MRSA has become a major concern.[90]

The spread of community-acquired MRSA clones is possibly related to the small size of the SCC*mec* types IV and V. There is a trade-off to the acquisition of resistance, which is that it imparts of fitness cost which may render the strain uncompetitive against susceptible strains when antibiotics are not present in the environment. This is thought to be the reason why infection, and carriage, of HA-MRSA clones have remained largely confined to healthcare settings. The smaller type IV and V SCC*mec* cassettes do not only confer multiple resistances, but may also result in a smaller fitness cost.

Although the epidemiological distinctions between CA-MRSA and HA-MRSA can be largely explained in terms of the fitness cost of resistance, the more general question of why a single MRSA clone can predominate in a given area, or the forces underlying clonal replacement, are far less well understood. It is probable that genetic differences underlie increased or decreased fitness (transmissibility),[64,75,76,91,92] and some general traits have been identified, which may account for epidemic spread. These include the ability to survive in the environment, to colonize the host, to multiply on epithelial and mucosal surfaces, to "detach" from the host, and to resist various antimicrobials. However, stochastic effects and extrinsic factors, such as local compliance to infection control measures and local use of antibiotics, may also have unpredictable consequences for the local composition of circulating MRSA clones. Furthermore, the specific genetic differences corresponding to fitness effects are very difficult to identify due to extensive gene redundancy and the possibility of subtle epistatic or regulatory effects playing a major role. The precise relationship between the "spread" (epidemicity) of a clone and its virulence potential is also unclear.

These complications can perhaps explain why a number of studies drawing comparisons between epidemic and sporadic MRSA have not generated clear experimental evidence consistent with the different epidemiological patterns.[93–99] An exception is a study demonstrating differences in biofilm production and adhesion to epithelial cells within epidemic variants of the Brazilian clone (ST239-III).[100] Although these laboratory comparisons were carried out on a small sample of strains, an epidemiological study also found evidence for increased virulence of an ST239 variant (TW20) that caused an outbreak in a London hospital.[101]

Molecular approaches have also not provided a clear understanding of epidemiological differences between clones. Population genetic analyses based on nucleotide sequence data of both housekeeping (MLST) genes and cell surface adhesion genes (which play a key role in host invasion) have also largely failed to detect robust links between genotype and epidemic phenotype.[39] Comparative genome hybridization and WGS have also been used to compare epidemic and sporadic strains, but this approach also failed to identify any genes likely to play a major role in increased transmission.[102,103] These findings are strong evidence against the presence or absence of a single common specific factor differentiating epidemic from sporadic *S. aureus* clones.

Although the evidence linking genotype and epidemiological phenotype is in many cases weak, there are tantalizing clues. For example, the CA-MRSA strain USA300 has disseminated widely throughout the United States. Genome sequencing of this strain

Downloaded From Vet-eBooks.com

revealed a novel genetic element, the arginine catabolic mobile element (ACME), which contained the gene for the arginine deiminase that may play a crucial role in the growth and survival of the bacterial cells.[53,104] However, studies reported in 2011 have shown that among 9−15% of the USA300 strains do not carry the ACME genomic region.[105] In addition, genome sequencing of 10 other isolates from the same disseminating clone confirmed its recent expansion.[106] Similarly, for the multidrug-resistant ST59 strains, a clone that is predominant in Taiwan has truncated *hsdM* and *hsdS* genes that encode the restriction−modification system. Hence, it was suggested that this deficiency in the restriction−modification system might have assisted the acquisition of mobile genetic elements from enterococci, which confer multidrug resistance.[107]

Besides being a human pathogen, *S. aureus* also colonizes the skin of animals and can cause a wide range of infections.[108−110] Livestock-associated MRSA (LA-MRSA) strains attract particular attention as the potential for zoonotic transmission raises the concern for public health. Previous reports have shown that distinct MRSA genotypes are associated with specific animal species. However, several studies have documented the transmission of LA-MRSA among different host species (e.g., from animals to humans and vice versa).[111−113] For instance, the LA-MRSA CC398 was first detected in pig farms and farmers from Europe, and has since been discovered to colonize and cause infections in other animals species (e.g., poultry, horses, dogs, and cattle) and humans worldwide.[114−118] Furthermore, CC398 was reported from humans lacking direct contact with livestock or livestock workers.[119−121] A study based on the whole-genome sequencing approach demonstrated that CC398 originated in humans as MSSA and was transmitted to livestock, where it subsequently acquired methicillin resistance.[122] Similarly, phylogenetic analysis of the MRSA CC5 poultry strains revealed that they have originated in humans and later transmitted to poultry, where they subsequently acquired avian-specific MGEs.[113] In contrast, it was shown in 2013 that the human pandemic MRSA CC97 strains recently made a bovine-to-human jump.[123] Taken together, these findings indicate that host-switches have been a feature in the evolution of a number of MRSA clones.

5. Conclusion

The widespread occurrence of MRSA in hospitals is recognized as a major challenge, especially with the emergence of strains with intermediate susceptibility to glycopeptides and of community-acquired MRSA. Given the difficulties to control MRSA, a thorough understanding of the processes underlying the emergence and spread of MRSA may help design new strategies to counteract this evolution. Several major pandemic clones have been identified and their epidemiology may change rapidly at a regional scale. Changes in clones have significant medical consequences, since the new clones often display different antibiotic susceptibility and/or virulence patterns.

The advances in sequencing technologies and the development of associated bioinformatics tools will provide a superior depth in the understanding of MRSA evolutionary history. These data will allow addressing many important questions about

Downloaded From Vet-eBooks.com

the evolution and epidemiology of MRSA and will bridge the gap left by the low discriminatory power of MLST. However, certain challenges concerning whole-genome sequencing still need to be addressed including choosing the proper strains collection, the development of standardized analysis pipeline, and the large-scale data management.

References

1. Skinner D, Keefer CS. Significance of bacteremia caused by *Staphylococcus aureus* − a study of one hundred and twenty-two cases and a review of the literature concerned with experimental infection in animals. *Archives Intern Med* 1941;**68**:851−75.
2. Lowy FD. Antimicrobial resistance: the example of *Staphylococcus aureus*. *J Clin Invest* 2003;**111**:1265−73.
3. Matsuhashi M, Song MD, Ishino F, et al. Molecular cloning of the gene of a penicillin-binding protein supposed to cause high resistance to beta-lactam antibiotics in *Staphylococcus aureus*. *J Bacteriol* 1986;**167**:975−80.
4. Katayama Y, Ito T, Hiramatsu K. A new class of genetic element, staphylococcus cassette chromosome mec, encodes methicillin resistance in *Staphylococcus aureus*. *Antimicrob Agents Chemother* 2000;**44**:1549−55.
5. Couto I, Wu SW, Tomasz A, de Lencastre H. Development of methicillin resistance in clinical isolates of *Staphylococcus sciuri* by transcriptional activation of the mecA homologue native to the species. *J Bacteriol* 2003;**185**:645−53.
6. Ibrahem S, Salmenlinna S, Virolainen A, et al. Carriage of methicillin-resistant staphylococci and their SCCmec types in a long-term-care facility. *J Clin Microbiol* 2009;**47**:32−7.
7. Chongtrakool P, Ito T, Ma XX, et al. Staphylococcal cassette chromosome mec (*SCCmec*) typing of methicillin-resistant *Staphylococcus aureus* strains isolated in 11 Asian countries: a proposal for a new nomenclature for *SCCmec* elements. *Antimicrob Agents Chemother* 2006;**50**:1001−12.
8. de Lencastre H, Oliveira D, Tomasz A. Antibiotic resistant *Staphylococcus aureus*: a paradigm of adaptive power. *Curr Opin Microbiol* 2007;**10**:428−35.
9. Hiramatsu K, Cui L, Kuroda M, Ito T. The emergence and evolution of methicillin-resistant *Staphylococcus aureus*. *Trends Microbiol* 2001;**9**:486−93.
10. Kuroda M, Ohta T, Uchiyama I, et al. Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*. *Lancet* 2001;**357**:1225−40.
11. Ito T, Hiramatsu K, Oliveira DC, et al. Classification of staphylococcal cassette chromosome mec (SCCmec): guidelines for reporting novel SCCmec elements. *Antimicrob Agents Chemother* 2009;**53**:4961−7.
12. Okon KO, Basset P, Uba A, et al. Cooccurrence of predominant Panton-Valentine leukocidin-positive sequence type (ST) 152 and multidrug-resistant ST 241 *Staphylococcus aureus* clones in Nigerian hospitals. *J Clin Microbiol* 2009;**47**:3000−3.
13. Hanssen AM, Sollid JUE. Multiple staphylococcal cassette chromosomes and allelic variants of cassette chromosome recombinases in *Staphylococcus aureus* and coagulase-negative stahylococci from Norway. *Antimicrob Agents Chemother* 2007;**51**:1671−7.
14. Miragaia M, Couto I, De Lancastre H. Genetic diversity among methicillin-resistant Staphylococcus epidermidis (MRSE). *Microb Drug Resistance-Mechanisms Epidemiol Dis* 2005;**11**:83−93.

Downloaded From Vet-eBooks.com

15. Kondo Y, Abe H, Jinmei H, et al. Multiple chemical ligation under thermal cycle. *Nucleic Acids Symp Ser (Oxf)* 2007:353−4.
16. Milheirico C, Oliveira DC, de Lencastre H. Update to the multiplex PCR strategy for assignment of mec element types in *Staphylococcus aureus*. *Antimicrob Agents Chemother* 2007;**51**:4537.
17. Oliveira DC, Milheirico C, Vinga S, de Lencastre H. Assessment of allelic variation in the ccrAB locus in methicillin-resistant *Staphylococcus aureus* clones. *J Antimicrob Chemother* 2006;**58**:23−30.
18. Oliveira DC, Tomasz A, de Lencastre H. Secrets of success of a human pathogen: molecular evolution of pandemic clones of meticillin-resistant *Staphylococcus aureus*. *Lancet Infect Dis* 2002;**2**:180−9.
19. Oliveira DC, de Lencastre H. Multiplex PCR strategy for rapid identification of structural types and variants of the mec element in methicillin-resistant *Staphylococcus aureus*. *Antimicrob Agents Chemother* 2002;**46**:2155−61.
20. Lina G, Durand G, Berchich C, et al. Staphylococcal chromosome cassette evolution in *Staphylococcus aureus* inferred from ccr gene complex sequence typing analysis. *ClinMicrobiolInfect* 2006;**12**:1175−84.
21. Garcia-Alvarez L, Holden MT, Lindsay H, et al. Meticillin-resistant *Staphylococcus aureus* with a novel mecA homologue in human and bovine populations in the UK and Denmark: a descriptive study. *Lancet Infect Dis* 2011;**11**:595−603.
22. Monecke S, Gavier-Widen D, Mattsson R, et al. Detection of mecC-positive *Staphylococcus aureus* (CC130-MRSA-XI) in diseased European hedgehogs (*Erinaceus europaeus*) in Sweden. *PLoS One* 2013;**8**:e66166.
23. Paterson GK, Harrison EM, Holmes MA. The emergence of mecC methicillin-resistant *Staphylococcus aureus*. *Trends Microbiol* 2014;**22**:42−7.
24. Basset P, Prod'hom G, Senn L, Greub G, Blanc DS. Very low prevalence of meticillin-resistant *Staphylococcus aureus* carrying the mecC gene in western Switzerland. *J Hosp Infect* 2013;**83**:257−9.
25. Kerschner H, Harrison EM, Hartl R, Holmes MA, Apfalter P. First report of mecC MRSA in human samples from Austria: molecular characteristics and clinical data. *New Microbes New Infect* 2015;**3**:4−9.
26. Paterson GK, Morgan FJ, Harrison EM, et al. Prevalence and characterization of human mecC methicillin-resistant *Staphylococcus aureus* isolates in England. *J Antimicrob Chemother* 2014;**69**:907−10.
27. Cuny C, Layer F, Strommenger B, Witte W. Rare occurrence of methicillin-resistant *Staphylococcus aureus* CC130 with a novel *mec*A homologue in humans in Germany. *PLoS One* 2011;**6**:e24360.
28. Harrison EM, Paterson GK, Holden MT, et al. Whole genome sequencing identifies zoonotic transmission of MRSA isolates with the novel mecA homologue mecC. *EMBO Mol Med* 2013;**5**:509−15.
29. Stegger M, Lindsay JA, Moodley A, Skov R, Broens EM, Guardabassi L. Rapid PCR detection of *Staphylococcus aureus* clonal complex 398 by targeting the restriction-modification system carrying sau1-hsdS1. *J Clin Microbiol* 2011;**49**:732−4.
30. Pichon B, Hill R, Laurent F, et al. Development of a real-time quadruplex PCR assay for simultaneous detection of nuc, Panton-Valentine leucocidin (PVL), mecA and homologue mecALGA251. *J Antimicrob Chemother* 2012;**67**.
31. Becker K, Denis O, Roisin S, et al. Detection of mecA- and mecC-positive methicillin-resistant *Staphylococcus aureus* (MRSA) isolates by the new Xpert MRSA gen 3 PCR assay. *J Clin Microbiol* 2016;**54**:180−4.

Downloaded From Vet-eBooks.com

32. Becker K, Larsen AR, Skov RL, et al. Evaluation of a modular multiplex-PCR methicillin-resistant *Staphylococcus aureus* detection assay adapted for mecC detection. *J Clin Microbiol* 2013;**51**:1917−9.
33. Enright MC, Robinson DA, Randle G, Feil EJ, Grundmann H, Spratt BG. The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). *Proc Natl Acad Sci USA* 2002;**99**:7687−92.
34. Nübel U, Roumagnac P, Feldkamp M, et al. Frequent emergence and limited geographic dispersal of methicillin-resistant *Staphylococcus aureus*. *Proc Natl Acad Sci USA* 2008; **105**:14130−5.
35. Feil EJ, Cooper JE, Grundmann H, et al. How clonal is *Staphylococcus aureus*? *J Bacteriol* 2003;**185**:3307−16.
36. Robinson DA, Enright MC. Evolutionary models of the emergence of methicillin-resistant *Staphylococcus aureus*. *Antimicrob Agents Chemother* 2003;**47**:3926−34.
37. Harris SR, Feil EJ, Holden MT, et al. Evolution of MRSA during hospital transmission and intercontinental spread. *Science* 2010;**327**:469−74.
38. Lindsay JA. Genomic variation and evolution of *Staphylococcus aureus*. *Int J Med Microbiol* 2010;**300**:98−103.
39. Kuhn G, Francioli P, Blanc DS. Evidence for clonal evolution among highly polymorphic genes in methicillin-resistant *Staphylococcus aureus*. *J Bacteriol* 2006;**188**:169−78.
40. Cooper JE, Feil EJ. The phylogeny of *Staphylococcus aureus* − which genes make the best intra-species markers? *Microbiology-Sgm* 2006;**152**:1297−305.
41. Hughes AL, Friedman R. Nucleotide substitution and recombination at orthologous loci in *Staphylococcus aureus*. *J Bacteriol* 2005;**187**:2698−704.
42. Sabat A, Wladyka B, Kosowska-Shick K, et al. Polymorphism, genetic exchange and intragenic recombination of the aureolysin gene among *Staphylococcus aureus* strains. *BMC Microbiol* 2008;**8**:129.
43. Narra HP, Ochman H. Of what use is sex to bacteria? *Curr Biol* 2006;**16**:R705−10.
44. Robinson DA, Enright MC. Evolution of *Staphylococcus aureus* by large chromosomal replacements. *J Bacteriol* 2004;**186**:1060−4.
45. Fitzgerald JR, Reid SD, Ruotsalainen E, et al. Genome diversification in *Staphylococcus aureus*: molecular evolution of a highly variable chromosomal region encoding the Staphylococcal exotoxin-like family of proteins. *InfectImmun* 2003;**71**:2827−38.
46. Robinson DA, Monk AB, Cooper JE, Feil EJ, Enright MC. Evolutionary genetics of the accessory gene regulator (agr) locus in *Staphylococcus aureus*. *J Bacteriol* 2005;**187**: 8312−21.
47. Grundmann H, Aanensen DM, van den Wijngaard CC, et al. Geographic distribution of *Staphylococcus aureus* causing invasive infections in Europe: a molecular-epidemiological analysis. *PLoS Med* 2010;**7**:e1000215.
48. Lan RT, Reeves PR. Intraspecies variation in bacterial genomes: the need for a species genome concept. *Trends Microbiol* 2000;**8**:396−401.
49. El Garch F, Hallin M, De Mendonca R, Denis O, Lefort A, Struelens MJ. StaphVar-DNA microarray analysis of accessory genome elements of community-acquired methicillin-resistant *Staphylococcus aureus*. *J Antimicrob Chemother* 2009;**63**:877−85.
50. Fitzgerald JR, Sturdevant DE, Mackie SM, Gill SR, Musser JM. Evolutionary genomics of *Staphylococcus aureus*: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. *Proc Natl Acad Sci USA* 2001;**98**:8821−6.
51. Lindsay JA, Holden MT. *Staphylococcus aureus*: superbug, super genome? *Trends Microbiol* 2004;**12**:378−85.

Downloaded From Vet-eBooks.com

52. Loffler B, Hussain M, Grundmeier M, et al. *Staphylococcus aureus* panton-valentine leukocidin is a very potent cytotoxic factor for human neutrophils. *PLoS Pathog* 2010; **6**:e1000715.
53. Highlander SK, Hulten KG, Qin X, Jiang H, Yerrapragada S, et al. Subtle genetic changes enhance virulence of methicillin resistant and sensitive *Staphylococcus aureus*. *BMC Microbiol* 2007;**7**:99.
54. Holden MT, Feil EJ, Lindsay JA, Peacock SJ, Day NP, et al. Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance. *Proc Natl Acad Sci USA* 2004;**101**:9786−91.
55. Ohta T, Hirakawa H, Morikawa K, et al. Nucleotide substitutions in *Staphylococcus aureus* strains, Mu50, Mu3, and N315. *DNA Res* 2004;**11**:51−6.
56. Sivaraman K, Venkataraman N, Tsai J, Dewell S, Cole AM. Genome sequencing and analysis reveals possible determinants of *Staphylococcus aureus* nasal carriage. *BMC Genomics* 2008;**9**:433.
57. Moore PC, Lindsay JA. Genetic variation among hospital isolates of methicillin-sensitive *Staphylococcus aureus*: evidence for horizontal transfer of virulence genes. *J Clin Microbiol* 2001;**39**:2760−7.
58. Peacock SJ, Moore CE, Justice A, et al. Virulent combinations of adhesin and toxin genes in natural populations of *Staphylococcus aureus*. *Infect Immun* 2002;**70**:4987−96.
59. Monecke S, Ehricht R, Slickers P, Wiese N, Jonas D. Intra-strain variability of methicillin-resistant *Staphylococcus aureus* strains ST228-MRSA-I and ST5-MRSA-II. *Eur J Clin Microbiol Infect Dis* 2009;**28**:1383−90.
60. Enright MC, Day NP, Davies CE, Peacock SJ, Spratt BG. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *J Clin Microbiol* 2000;**38**:1008−15.
61. Hallin M, Denis O, Deplano A, et al. Genetic relatedness between methicillin-susceptible and methicillin-resistant *Staphylococcus aureus*: results of a national survey. *J Antimicrob Chemother* 2007;**59**:465−72.
62. Aires de Sousa M, de Lencastre H. Bridges from hospitals to the laboratory: genetic portraits of methicillin-resistant *Staphylococcus aureus* clones. *FEMS Immunol Med Microbiol* 2004;**40**:101−11.
63. Blanc DS, Petignat C, Wenger A, et al. Changing molecular epidemiology of methicillin-resistant *Staphylococcus aureus* in a small geographic area over an eight-year period. *J Clin Microbiol* 2007;**45**:3729−36.
64. Aparicio P, Richardson J, Martin S, Vindel A, Marples RR, Cookson BD. An epidemic methicillin-resistant strain of *Staphylococcus aureus* in Spain. *Epidemiol Infect* 1992;**108**: 287−98.
65. de Lencastre H, Severina EP, Milch H, Konkoly Thege M, Tomasz A. Wide geographic disctribution of a unique methicillin-resistant *Staphylococcus aureus* clone in Hungarian hospitals. *Clin Microbiol Infect* 1997;**3**:289−96.
66. Deplano A, Witte W, van Leeuwen WJ, Brun Y, Struelens MJ. Clonal dissemination of epidemic methicillin-resistant *Staphylococcus aureus* in Belgium and neighboring countries. *Clin Microbiol Infect* 2000;**6**:239−45.
67. Gomes AR, Sanches IS, ires de SM, Castaneda E, de LH. Molecular epidemiology of methicillin-resistant *Staphylococcus aureus* in Colombian hospitals: dominance of a single unique multidrug-resistant clone. *Microb Drug Resist* 2001;**7**:23−32.
68. Leski T, Oliveira D, Trzcinski K, et al. Clonal distribution of methicillin-resistant *Staphylococcus aureus* in Poland. *J Clin Microbiol* 1998;**36**:3532−9.

Downloaded From Vet-eBooks.com

69. Melter O, Santos S,I, Schindler J, et al. Methicillin-resistant *Staphylococcus aureus* clonal types in the Czech Republic. *J Clin Microbiol* 1999;**37**:2798−803.
70. Oliveira D, Santos Sanches I, Mato R, et al. Virtuallay all methicillin-resistant *Staphylococcus aureus* (MRSA) infections in the largest Portuguese teaching hospital are caused by two internationally spread multiresistant strains: the "Iberian" and the "Brazilian" clones of MRSA. *Clin Microbiol Infect* 1998;**4**:373−84.
71. Aires de Sousa M, Sanches IS, van BA, van LW, Verbrugh H, de Lencastre H. Characterization of methicillin-resistant *Staphylococcus aureus* isolates from Portuguese hospitals by multiple genotyping methods. *Microb Drug Resist* 1996;**2**:331−41.
72. Ayliffe GA. The progressive intercontinental spread of methicillin-resistant *Staphylococcus aureus*. *Clin Infect Dis* 1997;**24**(Suppl. 1):S74−9.
73. Mato R, Sanches S, Venditti M, et al. Spread of the multiresistant Iberian clone of methicillin-resistant *Staphylococcus aureus* (MRSA) to Italy and Scotland. *Microb Drug Resist* 1998;**4**:107−12.
74. Oliveira DC, Tomasz A, de Lencastre H. The evolution of pandemic clones of methicillin-resistant *Staphylococcus aureus*: identification of two ancestral genetic backgrounds and the associated mec elements. *Microb Drug Resist* 2001;**7**:349−61.
75. Blanc DS, Petignat C, Moreillon P, et al. Unusual spread of a penicillin-susceptible methicillin-resistant *Staphylococcus aureus* clone in a geographic area of low incidence. *Clin Infect Dis* 1999;**29**:1512−8.
76. Cookson BD, Phillips I. Epidemic methicillin-resistant *Staphylococcus aureus*. *J Antimicrob Chemother* 1988;**21**(Suppl. C):57−65.
77. Perez-Roth E, Lorenzo-Diaz F, Batista N, Moreno A, Mendez-Alvarez S. Tracking methicillin-resistant *Staphylococcus aureus* clones during a 5-year period (1998 to 2002) in a Spanish hospital. *J Clin Microbiol* 2004;**42**:4649−56.
78. Amorim ML, ires de SM, Sanches IS, et al. Clonal and antibiotic resistance profiles of methicillin-resistant *Staphylococcus aureus* (MRSA) from a Portuguese hospital over time. *Microb Drug Resist* 2002;**8**:301−9.
79. Blanc DS, Francioli P, Le Coustumier A, et al. Reemergence of gentamicin-susceptible strains of methicillin-resistant *Staphylococcus aureus* in France: a phylogenetic approach. *J Clin Microbiol* 2001;**39**:2287−90.
80. Denis O, Deplano A, Nonhoff C, et al. National surveillance of methicillin-resistant *Staphylococcus aureus* in Belgian hospitals indicates rapid diversification of epidemic clones. *Antimicrob Agents Chemother* 2004;**48**:3625−9.
81. Pantazatou A, Papaparaskevas J, Stefanou I, Papanicolas J, Demertzi E, Avlamis A. Changes in the epidemiology of methicillin-resistant *Staphylococcus aureus* in a Greek tertiary care hospital, over an 8-year-period. *Int J Antimicrob Agents* 2003;**21**:542−6.
82. Rossney AS, Keane CT. Strain variation in the MRSA population over a 10-year period in one Dublin hospital. *Eur J Clin Microbiol Infect Dis* 2002;**21**:123−6.
83. Velazquez-Meza ME, ires de SM, Echaniz-Aviles G, et al. Surveillance of methicillin-resistant *Staphylococcus aureus* in a pediatric hospital in Mexico City during a 7-year period (1997 to 2003): clonal evolution and impact of infection control. *J Clin Microbiol* 2004;**42**:3877−80.
84. Feil EJ, Nickerson EK, Chantratita N, et al. Rapid detection of the pandemic methicillin-resistant *Staphylococcus aureus* clone ST 239, a dominant strain in Asian hospitals. *J Clin Microbiol* 2008;**46**:1520−2.
85. Chao G, Bao G, Jiao X. Molecular epidemiological characteristics and clonal genetic diversity of *Staphylococcus aureus* with different origins in China. *Foodborne Pathogens Dis* 2014;**11**:503−10.

Downloaded From Vet-eBooks.com

86. Dufour P, Gillet Y, Bes M, et al. Community-acquired methicillin-resistant *Staphylococcus aureus* infections in France: emergence of a single clone that produces Panton-Valentine leukocidin. *Clin Infect Dis* 2002;**35**:819−24.
87. Aires de Sousa M, Sanches IS, Ferro ML, et al. Intercontinental spread of a multidrug-resistant methicillin-resistant *Staphylococcus aureus* clone. *J Clin Microbiol* 1998;**36**: 2590−6.
88. Seybold U, Kourbatova EV, Johnson JG, et al. Emergence of community-associated methicillin-resistant *Staphylococcus aureus* USA300 genotype as a major cause of health care-associated blood stream infections. *Clin Infect Dis* 2006;**42**:647−56.
89. Patel M, Waites KB, Hoesley CJ, Stamm AM, Canupp KC, Moser SA. Emergence of USA300 MRSA in a tertiary medical centre: implications for epidemiological studies. *J Hosp Infect* 2008;**68**:208−13.
90. Skov RL, Jensen KS. Community-associated meticillin-resistant *Staphylococcus aureus* as a cause of hospital-acquired infections. *J Hosp Infect* 2009;**73**:364−70.
91. Marples RR, Cooke EM. Current problems with methicillin-resistant *Staphylococcus aureus*. *J Hosp Infect* 1988;**11**:381−92.
92. Witte W, Cuny C, Braulke C, Heuck D. Clonal dissemination of two MRSA strains in Germany. *Epidemiol Infect* 1994;**113**:67−73.
93. Duckworth GJ, Jordens JZ. Adherence and survival properties of an epidemic methicillin-resistant strains of *Staphylococcus aureus* compared with those of methicillin-sensitive strains. *J Med Microbiol* 1990;**32**:195−200.
94. Farrington M, Brenwald N, Haines D, Walpole E. Resistance to desiccation and skin fatty acids in outbreak strains of methicillin-resistant *Staphylococcus aureus*. *J Med Microbiol* 1992;**36**:56−60.
95. Jordens JZ, Duckworth GJ, Williams RJ. Production of "virulent factors" by "epidemic" methicillin-resistant *Staphylococcus aureus in vitro*. *J Med Microbiol* 1989;**30**:245−52.
96. Peacock JE, Moorman DR, Wenzel RP, Mandell GL. Methicillin-resistant *Staphylococcus aureus*: microbiologic characteristics, antimicrobial susceptibilities, and assessment of virulence of an epidemic strain. *J Infect Dis* 1981;**144**:575−82.
97. Roberts JIS, Gaston MA. Protein A and coagulase expression in epidemic and non-epidemic *Staphylococcus aureus*. *J Clin Pathology* 1987;**40**:837−40.
98. van Wamel WJB, Fluit AC, Wadström T, van Dijk H, Verhoef J, Vandenbroucke-Grauls CM. Phenotypic characterization of epidemic versus sporadic strains of methicillin-resistant *Staphylococcus aureus*. *J Clin Microbiol* 1995;**33**:1769−74.
99. Wagenvoort JHT, Penders RJR. Long-term in-vitro survival of an epidemic MRSA phage-group III-29 strain (letter to the Editor). *J Hosp Infect* 1997;**35**:322−5.
100. Amaral MM, Coelho LR, Flores RP, et al. The predominant variant of the Brazilian epidemic clonal complex of methicillin-resistant *Staphylococcus aureus* has an enhanced ability to produce biofilm and to adhere to and invade airway epithelial cells. *J Infect Dis* 2005;**192**:801−10.
101. Edgeworth JD, Yadegarfar G, Pathak S, et al. An outbreak in an intensive care unit of a strain of methicillin-resistant *Staphylococcus aureus* sequence type 239 associated with an increased rate of vascular access device-related bacteremia. *Clin Infect Dis* 2007;**44**: 493−501.
102. Kuhn G, Koessler T, Melles DC, et al. Comparative genomics of epidemic versus sporadic *Staphylococcus aureus* strains does not reveal molecular markers for epidemicity. *Infect Genet Evol* 2010;**10**:89−96.

Downloaded From Vet-eBooks.com

103. Vogel V, Falquet L, Calderon-Copete SP, Basset P, Blanc DS. Short term evolution of a highly transmissible methicillin-resistant *Staphylococcus aureus* clone (ST228) in a tertiary care hospital. *PLoS One* 2012;**7**:e38969.
104. Diep BA, Gill SR, Chang RF, et al. Complete genome sequence of USA300, an epidemic clone of community-acquired meticillin-resistant *Staphylococcus aureus*. *Lancet* 2006; **367**:731−9.
105. Uhlemann AC, Knox J, Miller M, et al. The environment as an unrecognized reservoir for community-associated methicillin resistant *Staphylococcus aureus* USA300: a case-control study. *PLoS One* 2011;**6**:e22407.
106. Kennedy AD, Otto M, Braughton KR, et al. Epidemic community-associated methicillin-resistant *Staphylococcus aureus*: recent clonal expansion and diversification. *Proc Natl Acad Sci USA* 2008;**105**:1327−32.
107. Hung WC, Takano T, Higuchi W, et al. Comparative genomics of community-acquired ST59 methicillin-resistant *Staphylococcus aureus* in Taiwan: novel mobile resistance structures with IS1216V. *PLoS One* 2012;**7**:e46987.
108. Witte W, Kresken M, Braulke C, Cuny C. Increasing incidence and widespread dissemination of methicillin-resistant *Staphylococcus aureus* in hospitals in central Europe, with special reference to German hospitals. *Clin Microbiol Infect* 1997;**3**:414−22.
109. Fitzgerald JR, Meaney WJ, Hartigan PJ, Smyth CJ, Kapur V. Fine-structure molecular epidemiological analysis of *Staphylococcus aureus* recovered from cows. *Epidemiol Infect* 1997;**119**:261−9.
110. Cuny C, Friedrich A, Kozytska S, et al. Emergence of methicillin-resistant *Staphylococcus aureus* (MRSA) in different animal species. *Int J Med Microbiol IJMM* 2010;**300**:109−17.
111. Strommenger B, Kehrenberg C, Kettlitz C, et al. Molecular characterization of methicillin-resistant *Staphylococcus aureus* strains from pet animals and their relationship to human isolates. *J Antimicrob Chemother* 2006;**57**:461−5.
112. van Duijkeren E, Wolfhagen MJ, Box AT, Heck ME, Wannet WJ, Fluit AC. Human-to-dog transmission of methicillin-resistant *Staphylococcus aureus*. *Emerg Infect Dis* 2004; **10**:2235−7.
113. Lowder BV, Guinane CM, Ben Zakour NL, et al. Recent human-to-poultry host jump, adaptation, and pandemic spread of *Staphylococcus aureus*. *Proc Natl Acad Sci USA* 2009; **106**:19545−50.
114. Witte W, Strommenger B, Stanek C, Cuny C. Methicillin-resistant *Staphylococcus aureus* ST398 in humans and animals, central Europe. *Emerg Infect Dis* 2007;**13**:255−8.
115. Rabello RF, Moreira BM, Lopes RM, Teixeira LM, Riley LW, Castro AC. Multilocus sequence typing of *Staphylococcus aureus* isolates recovered from cows with mastitis in Brazilian dairy herds. *J Med Microbiol* 2007;**56**:1505−11.
116. Vanderhaeghen W, Cerpentier T, Adriaensen C, Vicca J, Hermans K, Butaye P. Methicillin-resistant *Staphylococcus aureus* (MRSA) ST398 associated with clinical and subclinical mastitis in Belgian cows. *Veterinary Microbiol* 2010;**144**:166−71.
117. Loeffler A, Pfeiffer DU, Lloyd DH, Smith H, Soares-Magalhaes R, Lindsay JA. Meticillin-resistant *Staphylococcus aureus* carriage in UK veterinary staff and owners of infected pets: new risk groups. *J Hosp Infect* 2010;**74**:282−8.
118. van Cleef BA, Monnet DL, Voss A, et al. Livestock-associated methicillin-resistant *Staphylococcus aureus* in humans, Europe. *Emerg Infect Dis* 2011;**17**:502−5.
119. Cuny C, Nathaus R, Layer F, Strommenger B, Altmann D, Witte W. Nasal colonization of humans with methicillin-resistant *Staphylococcus aureus* (MRSA) CC398 with and without exposure to pigs. *PLoS One* 2009;**4**:e6800.

Downloaded From Vet-eBooks.com

120. van der Mee-Marquet N, Francois P, Domelier-Valentin AS, et al. Emergence of unusual bloodstream infections associated with pig-borne-like *Staphylococcus aureus* ST398 in France. *Clin Infect Dis* 2011;**52**:152−3.
121. Fan J, Shu M, Zhang G, et al. Biogeography and virulence of *Staphylococcus aureus*. *PLoS One* 2009;**4**:e6216.
122. Price LB, Stegger M, Hasman H, et al. *Staphylococcus aureus* CC398: host adaptation and emergence of methicillin resistance in livestock. *mBio* 2012:3.
123. Spoor LE, McAdam PR, Weinert LA, et al. Livestock origin for a human pandemic clone of community-associated methicillin-resistant *Staphylococcus aureus*. *mBio* 2013:4.
124. Blanc DS. The use of molecular typing for the epidemiological surveillance and investigation of endemic nosocomial infections. *Infect Genet Evol* 2004;**4**:193−7.
125. Murchan S, Kaufmann ME, Deplano A, et al. Harmonization of pulsed-field gel electrophoresis protocols for epidemiological typing of strains of methicillin-resistant *Staphylococcus aureus*: a single approach developed by consensus in 10 European laboratories and its application for tracing the spread of related strains. *J Clin Microbiol* 2003;**41**(4): 1574−85.
126. McDougal LK, Steward CD, Killgore GE, Chaitram JM, McAllister SK, Tenover FC. Pulsed-field gel electrophoresis typing of Oxacillin-resistant *Staphylococcus aureus* isolates from the United States: establishing a national database. *J Clin Microbiol* 2003;**41**: 5113−20.
127. Mulvey MR, Chui L, Ismail J, et al. Development of a Canadian standardized protocol for subtyping methicillin-resistant *Staphylococcus aureus* using pulsed-field gel electrophoresis. *J Clin Microbiol* 2001;**39**:3481−5.
128. Enright MC, Spratt BG. Multilocus sequence typing. *Trends Microbiol* 1999;**7**:482−7.
129. Hallin M, Deplano A, Denis O, de MR, de RR, Struelens MJ. Validation of pulsed-field gel electrophoresis and spa typing for long-term, nationwide epidemiological surveillance studies of *Staphylococcus aureus* infections. *J Clin Microbiol* 2007;**45**:127−33.
130. Koreen L, Ramaswamy SV, Graviss EA, Naidich S, Musser JM, Kreiswirth BN. *Spa* typing method for discriminating among *Staphylococcus aureus isolates*: implications for use of a single marker to detect genetic micro- and macrovariation. *J Clin Microbiol* 2004; **42**:792−9.
131. Shopsin B, Gomez M, Montgomery SO, et al. Evaluation of protein A gene polymorphic region DNA sequencing for typing of *Staphylococcus aureus* strains. *J Clin Microbiol* 1999;**37**:3556−63.
132. Basset P, Hammer NB, Kuhn G, Vogel V, Sakwinska O, Blanc DS. *Staphylococcus aureus* clfB and spa alleles of the repeat regions are segregated into major phylogenetic lineages. *Infect Genet Evol* 2009;**9**:941−7.
133. Kuhn G, Francioli P, Blanc DS. Double-locus sequence typing using *clfB* and *spa*, a fast and simple method for epidemiological typing of methicillin-resistant *Staphylococcus aureus*. *J Clin Microbiol* 2007;**45**:54−62.
134. Tang YW, Waddington MG, Smith DH, et al. Comparison of protein A gene sequencing with pulsed-field gel electrophoresis and epidemiologic data for molecular typing of methicillin-resistant *Staphylococcus aureus*. *J Clin Microbiol* 2000;**38**:1347−51.

Downloaded From Vet-eBooks.com

Origin and Emergence of HIV/AIDS

25

M. D'arc[1,2], L. Etienne[1,3], E. Delaporte[1], M. Peeters[1]
[1]UMI 233, Institut de Recherche pour le Développement (IRD), INSERM U1175 and Université de Montpellier, Montpellier, France; [2]Instituto Nacional de Câncer, Rio de Janeiro, Brazil; [3]International Center for Infectiology Research, INSERM U1111, Ecole Normale Supérieure de Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5308, 69364, Lyon, France

1. History of AIDS

AIDS or the acquired immune deficiency syndrome was first recognized between 1979 and 1981 among men having sex with men (MSM) in New York, Los Angeles, or San Francisco who presented with pneumonia caused by *Pneumocystiis carinii* and/or with symptoms of Kaposi sarcoma[1] (Fig. 25.1). Subsequently, patients with similar symptoms were seen among intravenous drug users (IDUs), hemophiliacs, Haitians, and Africans in Europe. In May 1983, the etiologic agent of AIDS, the human immunodeficiency virus (HIV), was identified.[2] In 1984, several authors reported AIDS cases in women and men in hospitals from sub-Saharan Africa, suggesting the existence of a heterosexual epidemic.[3–5] Seroepidemiological studies showed subsequently that a significant proportion of the population in certain regions of Africa was infected with HIV. In the early 1990s, the epidemic exploded in southern and eastern Africa, where in certain urban areas 25% of pregnant women were HIV positive.[6]

Molecular epidemiological studies revealed that the epicenter of the HIV pandemic is situated in central Africa, and more precisely the area of Kinshasa, the capital city of the Democratic Republic of Congo (DRC).[7,8] The virus had been introduced from Africa in Haiti in the 1960s before it started to circulate in North America (MRCA 1969) about a decade before the discovery and description of the first AIDS cases.[9,10] The last estimates show that today around 37 million people are infected with HIV (Fig. 25.1). More than 70% of HIV-infected persons live in sub-Saharan Africa. Increasing and earlier access to antiretroviral treatment (ART) during the 2000s improved life quality of HIV-infected individuals but reduced also the spread of HIV-1.[11] As a consequence, the number of new HIV infections dropped by 38% since 2001, nevertheless, still around 2.0 million people became infected in 2014. HIV/AIDS was one of the most important infectious diseases that emerged in the 20th century, and many efforts are still necessary to end HIV/AIDS as a public health threat in the 21st century as aimed by UNAIDS. It is thus important to understand where this virus came from, how it has been introduced into the human population, and which factors were associated with host adaptation and epidemic spread.

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00025-1
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

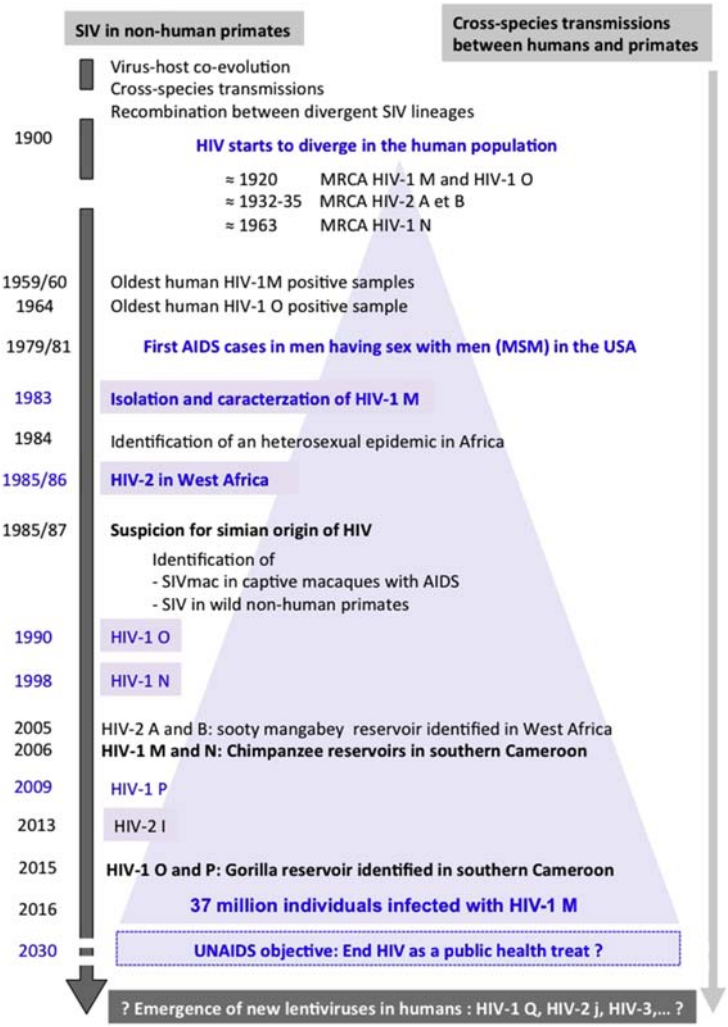


Figure 25.1 History of the AIDS epidemic: past and future events. Dates referring to events in the history of the HIV epidemic in humans are shown at the left, major events are highlighted in *blue* (*gray in print versions*), *boxes* indicate subsequent discoveries of the different HIV-1 groups. The number of persons living with HIV increases over time as illustrated by the *blue triangle*. *Gray arrows* (*dark gray in print versions*) represent a schematic time scale of the different events.

Downloaded From Vet-eBooks.com

2. Human Immunodeficiency Viruses Are Closely Related to Simian Immunodeficiency Virus From Nonhuman Primates

While HIV-1 has been identified in 1983, another closely related virus, HIV-2, has been described in 1985 in West Africa and among West Africans with AIDS in France.[12–14] HIV belongs to the *Lentivirus* genus of the Retroviridae family where five serogroups are recognized, each reflecting the vertebrate hosts with which they are associated (primates, sheep and goats, horses, cats, and cattle). Both HIV-1 and HIV-2 are most closely related to the lentiviruses from nonhuman primates (NHP), called simian immunodeficiency virus (SIV), and are thus most likely the result of cross-species transmissions of SIVs from African NHP.

2.1 Discovery of the First Simian Immunodeficiency Virus

Shortly after the identification of HIV-1 as the cause of AIDS in 1983, the first SIV, the SIVmac, was isolated from rhesus macaques (*Macaca mulatta*) with immune deficiency and clinical symptoms similar to AIDS at the New England Regional Primate Research Center (NERPRC) in the United States.[15,16] Retrospective studies revealed that SIVmac was introduced at NERPRC by other rhesus monkeys, previously housed at the California National Primate Research Center (CNPRC), where they survived an earlier disease outbreak (late 1960s), characterized by immune suppression and opportunistic infections.[17] A decade after the first outbreak, stump-tailed macaques (*Macaca arctoides*) developed a similar disease in the same settings, and a lentivirus, called SIVstm, was isolated from frozen tissue from one of these monkeys.[18] In both cases, the infected macaques had been in contact with healthy, but retrospectively shown, SIVsmm seropositive sooty mangabeys (*Cercocebys atys*) at the CNPRC.[18,19] The close phylogenetic relationship between SIVmac, SIVstm, and SIVsmm identified sooty mangabeys as the plausible source of SIV in captive macaques.[20]

Since SIVmac induced a disease in rhesus macaques with remarkable similarity to AIDS in humans, a simian origin of HIV was soon suspected. The discovery of HIV-2, the agent of AIDS in West Africa, and the remarkable relatedness of HIV-2 with SIVsmm, naturally occurring in sooty mangabeys in West Africa, reinforced this hypothesis.[21,22]

2.2 Simian Immunodeficiency Viruses in African Non-human Primates

Currently, serological evidence of SIV infection has been shown in at least 45 different NHP species, and the infection has been confirmed by viral sequence analysis in the majority of them (Table 25.1 and Fig. 25.2).[23] A high genetic diversity is observed among the different SIVs, but generally, each NHP species is infected with a species-specific virus, which forms monophyletic lineages in phylogenetic trees. These species-specific SIVs are identified by a lower-case-three-letter code, which

Downloaded From Vet-eBooks.com

Table 25.1 Simian Immunodeficiency Virus Infections in Old World Monkeys and Apes in Africa

Genus	Species/Subspecies	Common Name	SIV
Pan	***troglodytes troglodyte***s	***Central chimpanzee***	**SIVcpz*Ptt***
	troglodytes schweinfurthii	Eastern chimpanzee	SIVcpz*Pts*
Gorilla	***gorilla gorilla***	***Western lowland gorilla***	**SIVgor**
Colobus	*guereza*	Mantled guereza	*SIVcol-1*, *col-2*
Piliocolobus	*badius badius*	Western red colobus	SIVwrc*Pbb*
	badius temminckii	Temminck's red colobus	SIVwrc*Pbt*
	tholloni	Tshuapa red colobus	SIVtrc[a]
	rufomitratus tephrosceles	Ugandan red colobus	SIVkrc
Procolobus	*verus*	Olive colobus	SIVolc
Lophocebus	*albigena*	Gray-cheeked managabey	[b]
	aterrimus	Black crested mangabey	SIVbkm[a]
Papi	*anubis*	Olive baboon	[b]
	cynocephalus	Yellow baboon	SIVagm-Ver[a]
	ursinus	Chacma baboon	SIVagm-Ver[a]
Cercocebus	***atys***	***Sooty mangabey***	**SIVsmm**
	torquatus	Red-capped mangabey	SIVrcm
	agilis	Agile mangabey	SIVagi
Mandrillus	*sphinx*	Mandrill	SIVmnd-1, mnd-2
	leucophaeus	Drill	SIVdrl
Allenopithecus	*nigroviridis*	Allen's swamp monkey	SIVasm[a,c]
Miopithecus	*talapoin*	Angolan talapoin	SIVtal[a]
	ogouensis	Gabon talapoin	SIVtal
Erythrocebus	*patas*	Patas monkey	SIVagm-Sab[a]
Chlorocebus	*sabaeus*	Green monkey	SIVagm-Sab

Downloaded From Vet-eBooks.com

Table 25.1 Simian Immunodeficiency Virus Infections in Old World Monkeys and Apes in Africa—cont'd

Genus	Species/Subspecies	Common Name	SIV
	aethiops	Grivet	SIVagm-Gri
	tantalus	Tantalus monkey	SIVagm-Tan
	pygerythrus	Vervet monkey	SIVagm-Ver
Cercopithecus	*diana*	Diana monkey	[b]
	nictitans	Greater spot-nosed monkey	SIVgsn
	mitis	Blue monkey	SIVblu[a]
	albogularis	Sykes's monkey	SIVsyk
	mona	Mona monkey	SIVmon
	campbelli	Campbell's mona	[b]
	pogonias	Crested mona	[b]
	denti	Dent's mona	SIVden
	cephus	Mustached guenon	SIVmus-1,mus-2, mus-3
	erythrotis	Red-eared monkey	SIVery[a]
	ascanius	Red-tailed monkey	SIVasc
	lhoest	l'Hoest monkey	SIVlho
	solatus	Sun-tailed monkey	SIVsun
	preussi	Preuss's monkey	SIVpre[a]
	hamlyni	Owl-faced monkey	[b]
	neglectus	de Brazza's monkey	SIVdeb
	wolfi	Wolf's monkey	SIVwol[a]

For each species, the genus, species, and subspecies (if applicable) are given. Species representing a reservoir for HIV-1 and -2 are highlighted in bold.
[a]Only partial SIV sequences are available.
[b]Only serological evidence for SIV.
[c]Unpublished data from Ahuka-Mundeke S, et al.

corresponds to the initial letters of the common species name; for example, SIVagm for African green monkeys and SIVrcm for red-capped mangabeys. When different subspecies of the same species are infected, an abbreviation referring to the name of the subspecies is added to the virus designation, that is, SIVcpz*Ptt* and SIVcpz*Pts* to differentiate between the two chimpanzee subspecies, *Pan troglodytes troglodytes* and *P. t. schweinfurthii*, respectively.

Downloaded From Vet-eBooks.com

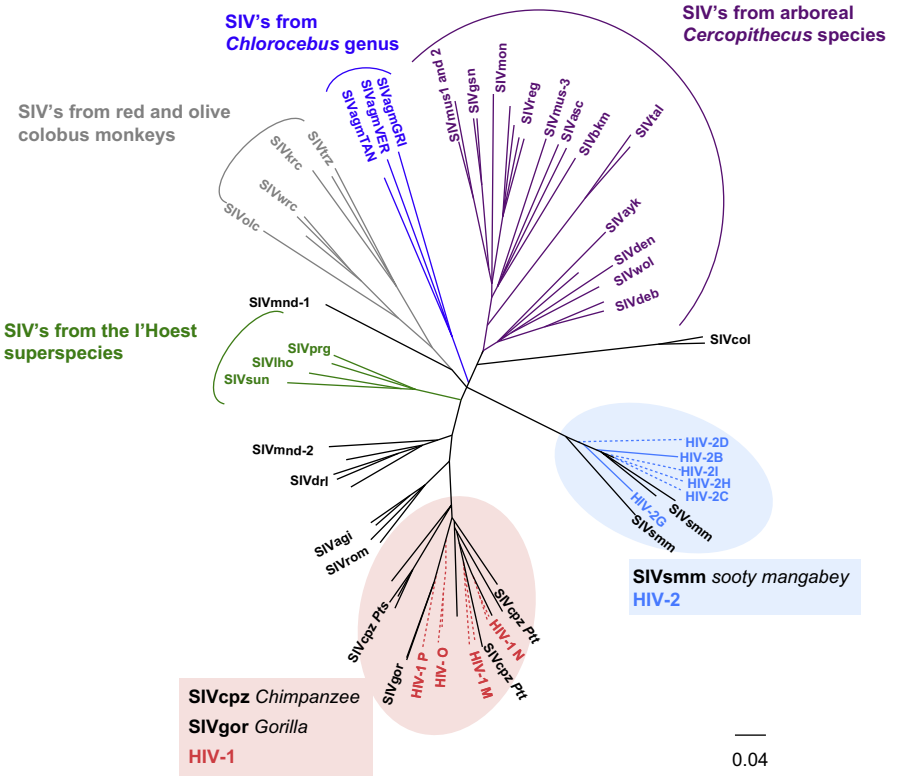


Figure 25.2 Genetic diversity and evolutionary history of the different HIV/SIV lineages. Phylogenetic tree analysis using the neighbor-joining method on a 294 bp fragment from the *pol* gene of different SIVs infecting non-human primates and HIVs infecting humans. Branch lengths are drawn to scale (the scale bar indicates 0.04 substitutions per site). The different colors are used for clarity to discriminate the different SIV lineages. The different HIV-1 (*red-gray in print versions*) and HIV-2 (*blue-gray in print versions*) lineages are interspersed with the SIVcpz/SIVgor (*red circle*) and SIVsmm (*blue circle*) lineages, respectively. The correspondence between the SIV lineages and their natural hosts are shown in Table 25.1.

Importantly, the number of African NHP infected with SIV is most probably underestimated, since around 30 species of the 73 actually recognized Old World NHP species in sub-Saharan Africa have not been tested yet or perhaps only very few individuals have been tested. Knowing that the vast majority (90%) of NHP species tested is infected with SIV, many of the remaining species can be expected to harbor SIV infections. In addition, the majority of SIV lineages known to date were discovered by HIV screening or confirmatory assays, which are based on cross-reactivity with HIV-1 and/or HIV-2 antigens. Because the extent of this cross-reactivity is not known, SIV infection can be underestimated. To increase sensitivity and chances to uncover new SIV lineages, Elisas using SIV-specific antigens were developed, and the multiple analyte profiling (xMAP) technology has

Downloaded From Vet-eBooks.com

also been adapted to allow a single sample to be tested simultaneously against antigens from a wide diversity of SIV lineages.[24–27] This latter approach reduces workload and limits the amount of biological material, which is often only available in very low quantities.[27] Using these SIV lineage-specific antibody assays, new SIV lineages have been identified, and it has been shown that prevalence rates can differ significantly not only among species (0% to >40%), but also within species according to the sampling site.[26,27]

Interestingly, only Old World primates are infected with SIVs, and only those from the African continent. No SIVs have been identified in Asian primate species, but it has to be noted that studies on Asian and New World primates are scarce.[28,29]

The widespread presence of SIVs in numerous African NHP suggests that SIV is very old. A report studied SIVs in NHP from Bioko Island, Equatorial Guinea, and related species on the continent suggest that SIVs evolved independently since Bioko became isolated from the African mainland 10,000–12,000 years ago with the elevation of the sea levels. Molecular clock analysis that used the date of the separation of Bioko Island to calibrate showed that SIVs have been present in African primates for more than 32,000 years.[30] More recently, a study reported in 2013 on SIVagm-Ver diversity in vervets on the two sides of the Drakensberg Mountain in South Africa estimated that the origin of SIVs could be retraced to 800,000–2,500,000 years.[31]

2.3 Pathogenicity of Simian Immunodeficiency Viruses in Their Natural Hosts

Although SIVs are called immune deficiency viruses, these viruses generally do not induce an AIDS-like disease in their natural hosts, suggesting that they evolved with their hosts over an extended period of time.[32] This absence of pathogenicity has been extensively studied not only in captive but also in naturally infected sooty mangabeys and African green monkeys. Their life span, as well as their immunological system, do not seem to be affected by SIV, despite the presence of high viral loads in blood and tissues.[33–35] However, some cases of AIDS have been described in captive monkeys, but mainly at an age that is not reached in their natural habitats, for example, two mandrills developed AIDS, after 17 years of SIVmnd infection and a sooty mangabey progressed to AIDS after 18 years.[36] Nevertheless, the paradigm of nonpathogenicity has been challenged in 2009 by observations on wild chimpanzees (*P. t. schweinfurthii*) from the Gombe National Park in Tanzania.[33] SIVcpz*Pts* infection was associated with a 10–16-fold age-corrected increased risk of mortality and reduced fertility in terms of an average number of births and survival of newborns among infected females. Moreover, retrospective analysis of tissues conserved from dead SIVcpz*Pts*-infected animals showed that they have also symptoms of immune deficiency. Similarly, a report on a young naturally SIV-infected *P. t. troglodytes* chimpanzee confiscated in Cameroon in 2003, subspecies in which the ancestors of HIV-1 have been documented, also suggests clinical progression to an AIDS-like disease. Clinical follow-up and biological analysis over a 7-year period showed a significant decline of CD4 counts, severe

Downloaded From Vet-eBooks.com

thrombocytopenia, weight loss, and frequent periods of infections with diverse pathogens. Finally, the animal died at the age of 10 years.[37]

An in-depth demographic and epidemiological study of the habituated chimpanzee populations (13% SIVcpz infected) from Gombe National Park together with those from a neighboring unhabituated community (Kalande) showed a catastrophic population decline in the highly SIVcpz-infected (46%) Kalande chimpanzee population.[38] Mathematical models applied on these chimpanzee populations showed that SIVcpz infection can cause serious population decline. Models also showed that depending on the population structure and transmission dynamics, SIVcpz may be more likely to go extinct than its chimpanzee host, and intercommunity migration increases survival of infected populations. These results are in line with the unequal distribution of SIVcpz in Africa and may explain how chimpanzees as a species have survived this pathogen.

The fact that such a major effect of SIVcpz infection went undetected for decades supports the need for more studies on the natural history of SIV infection in African NHP in their native habitat, and not in captive environments where health status is controlled, nutrition is monitored, and exposure to infectious agents is limited.

2.4 Evolution and Phylogeny of Simian Immunodeficiency Viruses

The genetic diversity among NHP lentiviruses is extremely complex. There are many examples of coevolution between viruses and their hosts, but also recombination between distant SIVs is not exceptional. Although it now seems clear that a simple codivergence between viruses and their hosts is not common, phylogenies for some SIVs and their hosts suggest coevolution over long time periods (Fig. 25.2). This is the case for the four different African green monkey species from the *Chlorocebus* genus that live in geographically separate and nonoverlapping areas across Africa. Each *Chlorocebus* species is infected with a specific SIV, for example, SIVagm-Ver in vervets, SIVagm-Gri in grivets, and SIVagm-Tan and SIVagm-Sab in tantalus and sabaeus monkeys, respectively.[39] The SIVs from the l'hoesti superspecies, (e.g., SIVlho from *Cercopithecus lhoesti*, SIVsun from *Cercopithecus solatus*, and SIVprg from *Cercopithecus preussi*) and SIVs from arboreal *Cercopithecus* species each also form separate clusters in the phylogenetic tree of SIVs.[40,41] However, the characterization of a new SIVagm-Tan strain form a tantalus monkey in Cameroon reported in 2015 showed a mosaic structure between SIVagm-Sab from sabaeus monkeys from West Africa and SIVagm-Tan from tantalus monkeys from Central Africa suggesting that the evolution of SIVagm in the *Chlorocebus* genus is more complex than previously thought.[42]

Nevertheless, cross-species transmissions could also give erroneous impressions of coevolution, especially when chances for efficient host switch are higher among genetically closely related species.[43] There are indeed numerous examples of cross-species transmissions of SIVs between primates living in the same habitats or in polyspecific associations. For example, SIVagm from African green monkeys has been transmitted

Downloaded From Vet-eBooks.com

to Patas monkeys in West Africa and to yellow and chacma baboons in South Africa.[44–46] There are also more complex examples of cross-species transmissions of SIVs between greater spot-nosed monkeys (SIVgsn) and mustached monkeys (SIVmus), followed by recombination as seen for SIVmus-2 in mustached monkeys in Cameroon or SIVmus-3 in Gabon[47,48] (Table 25.1). One of the most striking examples of cross-species transmission, followed by recombination is SIVcpz in chimpanzees. The 5′ region of SIVcpz is most similar to SIVrcm from red-capped mangabeys, and the 3′ region is found to be closely related to SIVgsn from greater spot-nosed monkeys.[49] Chimpanzees are known to hunt monkeys for food. Most probably, the recombination of these monkey viruses occurred within chimpanzees and gave rise to the common ancestor of today's SIVcpz lineages, which in turn were subsequently transmitted to gorillas and humans.[50–52] Some NHP are infected with more than one SIV lineage, often as a result of cross-species transmission and recombination, for example, SIVmnd-1 in mandrills from southern Gabon, and SIVmnd-2 in animals living in northern Gabon and Cameroon or SIVmus in mustached monkeys in which three different variants have been described.[48,53]

As numerous cross-species transmissions among different primate species occurred, both HIV-1 and HIV-2 in humans are also the results of cross-species transmissions of SIVs from African primates.[54] The closest simian relatives of HIV-1 are SIVcpz and SIVgor, in chimpanzees (*Pan troglodytes troglodytes*) and gorillas (*Gorilla gorilla gorilla*), respectively, from West Central Africa, and SIVsmm in sooty mangabeys (*Cercocebus atys*) from West Africa are the closest relatives of HIV-2.

3. HIV-1 Is Derived From Simian Immunodeficiency Viruses Circulating Among African Apes

Based on phylogenetic analyses of numerous isolates obtained from diverse geographic origins, HIV-1 is classified into four groups, M, N, O, and P (Fig. 25.1). Group M (for Major), discovered in 1983,[2] represents the vast majority of HIV-1 strains found worldwide and is responsible for the global pandemic. Group O (for Outlier), described in 1990, remained restricted to West Central Africa, and represents currently less than 1% of HIV-1 infections in Cameroon, and is estimated to have infected a cumulative number of 100,000 individuals.[55–58] Group N and group P, described in 1998 and 2009, respectively, have only been observed in a handful of patients, all from Cameroon except one HIV-1 N case.[59,60] Each HIV-1 group corresponds to an independent cross-species transmission of SIVs from apes to humans.

3.1 Simian Immunodeficiency Viruses From Chimpanzees and Gorillas Are the Ancestors of HIV-1 in Humans

The first SIVcpz strains have been isolated from two captive but wild-born chimpanzees in Gabon more than 25 years ago, SIVcpzGab1 and SIVcpzGab2.[61] Genetic analysis of the SIVcpzGab1 genome revealed the presence of the accessory gene, *vpu*, also

Downloaded From Vet-eBooks.com

identified in HIV-1. Furthermore, phylogenetic analysis showed that SIVcpzGab1 was more closely related to HIV-1 than to any other SIV.[62] Characterization of a third SIVcpz, SIVcpzANT, showed an unexpected high degree of divergence among the chimpanzee viruses.[63,64] Subsequent subspecies identification of the chimpanzee hosts revealed that the SIVcpzANT strain was isolated from a member of the *P. t. schweinfurtii* subspecies, whereas the other chimpanzees belonged to the *P. t. troglodytes* subspecies.[65] These findings suggested two distinct SIVcpz lineages according to the host subspecies: SIVcpz*Ptt* and SIVcpz*Pts* from central (*P. t. troglodytes*) and eastern chimpanzees (*P. t. schweinfurtii*), respectively. All HIV-1 groups were more closely related to SIVcpz*Ptt* than to SIVcpz*Pts* (Fig. 25.3). Although these data pointed already to the West Central African chimpanzees (*P. t. troglodytes*) as the natural reservoir of the ancestors of HIV-1, the SIVcpz reservoirs in wild-living apes that are at the origin of HIV in humans still needed to be identified. The major issue in studying SIVcpz infection in wild chimpanzees is their endangered status and the fact that they live in isolated forest regions. In addition, all previously studied chimpanzees

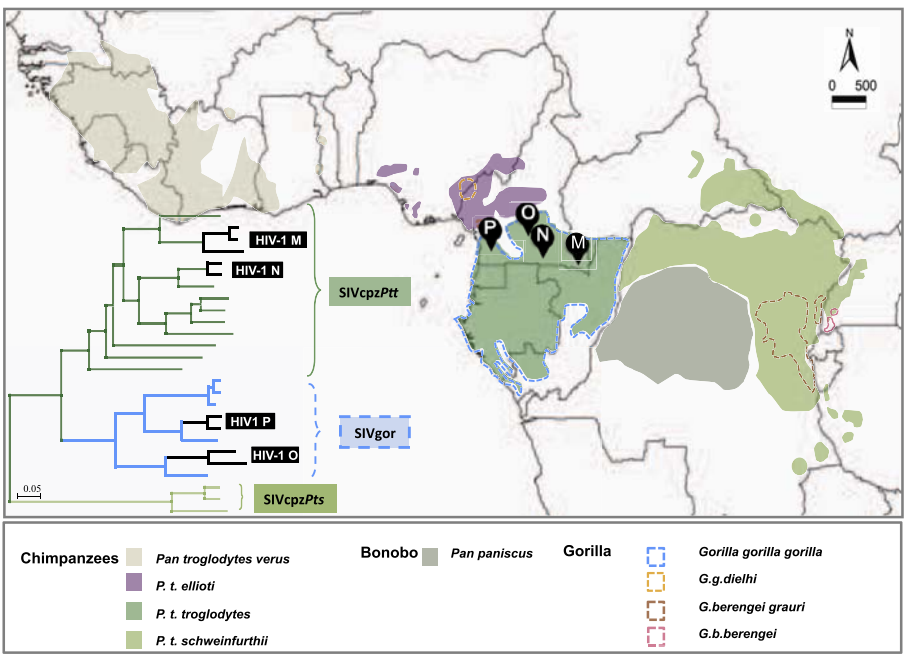


Figure 25.3 HIV-1 is derived from SIVs circulating in chimpanzees and gorillas from West Central Africa. Evolutionary relationship of SIVcpz*Pts* (*olive green-dark gray in print versions*), SIVcpz*Ptt* (*green-gray in print versions*), SIVgor (*blue-light gray in print versions*) and HIV-1 group M, N, O and P (*black*) strains based on maximum likelihood phylogenetic analysis of partial *env* (gp41) sequences. Horizontal branch lengths are drawn to scale (the scale bar indicates 0.05 substitutions per site). The map represents the geographical range of the four gorilla and the four chimpanzee subspecies, and bonobos. Arrows on the map indicate the ape populations infected with the ancestors of each HIV-1 group.

Downloaded From Vet-eBooks.com

were wild-caught, but they were captured as infants, which do not reflect true prevalences among wild living adult animals. The development of noninvasive methods to detect and characterize SIVcpz in fecal and urine samples boosted the search for new SIVcpz strains in wild ape populations in Africa.[66,67] The first full-length SIVcpz sequence from a wild-infected chimpanzee, SIVcpzTan1, was obtained from a fecal sample from a *P. t. schweinfurtii* chimpanzee in Gombe National Park, Tanzania.[68]

Subsequently, large-scale studies have been conducted across West, Central, and East Africa, and today more than 6000 fecal samples have been collected from the four different chimpanzee subspecies. These studies showed that only the two subspecies from Central Africa are infected and also confirmed the presence of subspecies-specific SIVcpz*Ptt* and SIVcpz*Pts* lineages in *P. t. troglodytes* and *P. t. schweinfurthii*, respectively. Interestingly, phylogeographic clusters are also observed within the SIVcpz*Ptt* and SIVcpz*Pts* lineages. As such, the reservoirs of the ancestors of the pandemic HIV-1 group M were identified in chimpanzee communities in southeast Cameroon. Similarly, the ancestors of HIV-1 group N have been identified in chimpanzees living in South-Central Cameroon. The overall SIVcpz prevalence ranges between 10% and 13% in both chimpanzee subspecies, but they are very heterogeneous and can vary from 0% to >40% in certain chimpanzee communities.[69–73] Only SIVcpz*Ptt* strains have been transmitted to humans, despite the fact that both chimpanzee subspecies represent significant SIVcpz reservoirs. Although numerous samples have been tested, no SIV infection has been detected in the two other chimpanzee subspecies, *P. t. elioti* (previously *P. t. vellerosus*) and *P. t. verus*.[72,74,75]

While the reservoirs of the ancestors of HIV-1 M and N were identified in 2006, the reservoirs of group O and P remained unknown. In 2006, SIV infection was described for the first time in western gorillas (*Gorilla gorilla gorilla*) in Cameroon. Surprisingly, the newly characterized gorilla viruses, termed SIVgor, formed a monophyletic group within the HIV-1/SIVcpz*Ptt* radiation, but in contrast to SIVcpz*Ptt*, they were most closely related to HIV-1 group O and P[51,52] (Fig. 25.3). Moreover, phylogenetic relationships between SIVcpz, SIVgor, and HIV-1 indicate that chimpanzees represent the original reservoir of SIVs now found in chimpanzees, gorillas, and humans (Fig. 25.3). However, it is not clear yet how the virus has been transmitted from chimpanzees to gorillas because physical encounters between the two species, such as biting or other contact with infected blood or body fluids, have not been reported and thus occur most likely very rarely. However, chimpanzees and gorillas have overlapping habitats and often feed in the same fruit trees,[76–78] which leads to direct and indirect contacts, and has also resulted in the cross-species transmission of other pathogens, for example, ebola and hepatitis B.[79,80]

Today almost 6000 fecal samples have been studied from all gorilla species and subspecies, for example, western lowland (*Gorilla gorilla gorilla*) and cross river (*G. g. diehli*) gorillas in West-Central Africa, and eastern lowland (*G. beringei graueri*) and mountain (*G. b. beringei*) gorillas in East-Central Africa.[50,81] Among the 55 different sites studied across Central Africa, covering a large proportion of the geographic range of gorillas, SIVgor infection were only identified in six sites all located in southern Cameroon and only in western lowland gorillas.[50,81] The global prevalence of SIVgor (<2%) is lower than the prevalence observed in SIVcpz

Downloaded From Vet-eBooks.com

infection from chimpanzees. Similarly, as for chimpanzees, an unequal distribution is observed, ranging from 0% to >30%.[50,81] Despite the low prevalence, the genetic characterization of SIVgor strains from different field locations showed a high genetic diversity and phylogeographic clustering. As such, HIV-1 group P viruses fall within the radiation of the SIVgor strains from a gorilla community in Southwest Cameroon, strongly suggesting the origin of HIV-1 P in this region of western Cameroon. Finally, sequence analysis of an SIVgor strain from South-Central Cameroon resolved the origin of HIV-1 group O.[50]

It is now clear that central chimpanzees (*P. t. troglodytes*) are the reservoirs for the pandemic HIV-1 group M strain and HIV-1 N, and western lowland gorillas (*G. g. gorilla*) are the reservoirs for HIV-1 groups O and P. The origin of all HIV-1 groups has now been solved and today we know that two ape species are involved in the origin of HIV-1 in humans, each species transmitted SIV to humans on at least two occasions.

Importantly, the finding that SIVcpz strains from East African chimpanzees, including those from Kisangani in DRC, are more distantly related to HIV-1 also provides evidence that the Oral Polio Vaccine (OPV), which was largely distributed in this part of Africa at the end of the 50 s, is not at the origin of the HIV-1 epidemic.[73] It has been suggested that tissues derived from SIVcpz-infected chimpanzees, captured in the northeastern part of DRC, were used for the OPV production. However, this geographical region is situated in the middle of the *P. t. schweinfurtii* range and the characterization of SIVcpz*Pts* from wild chimpanzees in DRC proved once more the inconsistency of the OPV theory[71,73] (Fig. 25.3).

3.2 The Cross-Species Transmissions Resulting in HIV-1 Viruses in Humans Occurred in West–Central Africa

Since the four groups of HIV-1 fall within the HIV-1/SIVcpz*Ptt*/SIVgor radiation, the cross-species transmissions giving rise to HIV-1 occurred in western equatorial Africa, the home of central chimpanzees (*P. t. troglodytes*) and western lowland gorillas (*G. g. gorilla*). Furthermore, no human counterpart is found for SIVcpz*Pts* from *P. t. schweinfurtii*. The studies in wild chimpanzee and gorilla communities in Cameroon, not only strengthen the evidence of the West Central African origin of HIV-1, but also indicate that the four groups of HIV-1 arose from geographically distinct chimpanzee and gorilla populations in southern Cameroon (Fig. 25.3). This coincides with the geographical area of group N, O, and P infections, which remain actually mainly restricted to Cameroon.[60,82,83]

The four HIV-1 groups thus have their origin in West Central Africa, but only one, HIV-1 group M, has spread across Africa and all the other continents. Moreover, the reservoir of the ancestors of HIV-1 M has been identified at almost 1000-km distance from the epicenter of the HIV-1 epidemic in DRC.[7,8] A combination of several factors (viral, host, socioeconomic, demographic, etc.) are thus most likely involved in the subsequent efficient spread of HIV-1 M.

Downloaded From Vet-eBooks.com

3.3 HIV-1 Started to Diverge in the Human Population at the Beginning of the 20th Century

As mentioned earlier, the SIVcpz*Ptt* ancestors of group M have been identified in Cameroon, but the highest genetic diversity of HIV-1 M, in number of cocirculating subtypes and intrasubtype diversity, has been observed in the western part of DRC, suggesting that this region may be the epicenter of HIV-1 group M.[7] Retrospective studies showed that 20 years before, the AIDS epidemic was recognized in the United States, HIV-1 M (subtypes A and D) infection was already circulating in humans in Kinshasa; for example, HIV-1 was identified in a serum from 1959 and a biopsy from 1960.[8,84] Similarly, HIV-1 group O was identified in a Norwegian sailor who became infected in the 1960s.[85] The most recent common ancestors of HIV-1 groups M and O are both estimated around 1920.[86,87] The low spread and lower levels of diversity seen in group N and especially in group P indicate that they most likely emerged more recently. The most recent common ancestor of HIV-1 group N is estimated around 1963 with a confidence interval of 1948−1977.[87] Only two HIV-1 P-infected patients have been identified today, and estimates on dates are uncertain. Coalescent studies suggest that groups O and M underwent similar rates of exponential growth until about 1960, after which group M undergoes accelerated epidemic growth.[86]

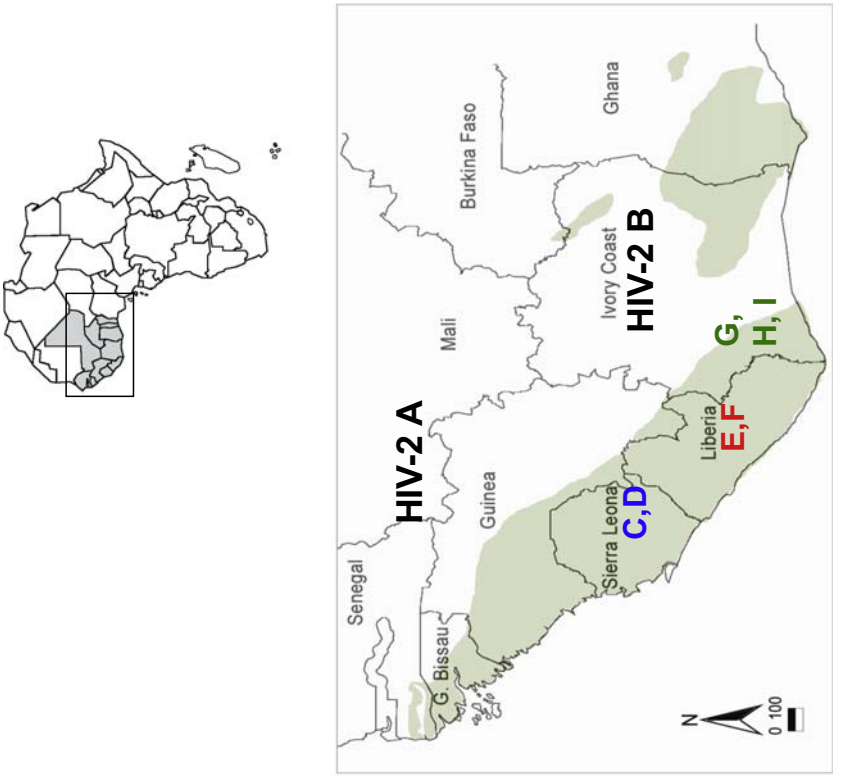
3.4 Simian Immunodeficiency Viruses Are Transmitted to Humans by Exposure to Infected Primates

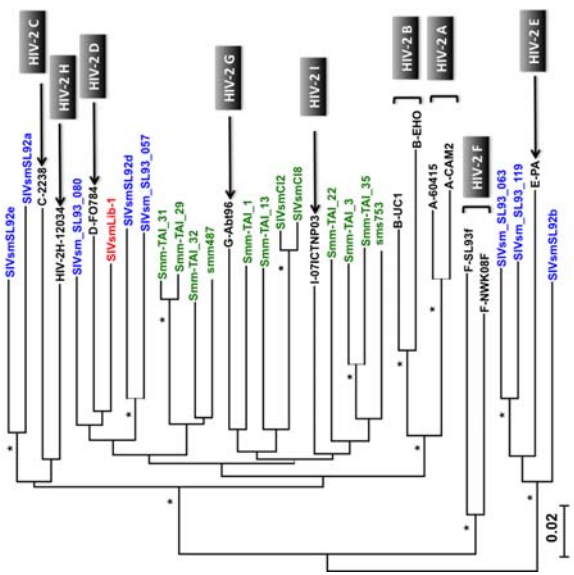
Although the conditions and circumstances of cross-species from SIVs from primates to humans remain unknown, human exposure to blood or other secretions of infected primates, through hunting and butchering of primate bushmeat, represents the most plausible source for human infection. In addition, bites and other injuries caused by primates kept as pet animals can favor a possible viral transmission. However, factors associated with single cross-species transmission have to be differentiated from those associated with the epidemic spread, the latter being a combination of viral, host, and environmental factors.

4. Origin of HIV-2: Another Emergence, Another Epidemic

Two independent studies in 1989 and 1992 confirmed the homologies between HIV-2 and SIVsmm infecting sooty mangabeys in West Africa.[21,22] Sooty mangabeys are home to West Africa, from Senegal to Ivory Coast, coinciding with the endemic region of HIV-2 (Fig. 25.4). In contrast to HIV-1, HIV-2 remained restricted to West Africa and HIV-2 prevalences are even decreasing, since HIV-1 M is now also predominating in West Africa.[88] The highest HIV-2 prevalences have been observed in Guinea-Bissau and southern Senegal (Casamance area). Overall, HIV-2 is less pathogenic,

Downloaded From Vet-eBooks.com





Downloaded From Vet-eBooks.com

less transmissible with almost absence of mother to child transmission, and less efficient in sexual transmission most likely related to lower viral loads.[89,90] However, a high genetic diversity is seen among HIV-2 strains, nine groups (A–I) of HIV-2 have been described so far[91,92] (Fig. 25.4), each corresponding to a cross-species transmission. Only groups A and B are largely represented in the HIV-2 epidemic, with group A circulating in the western part of West Africa (Senegal, Guinea-Bissau) and group B being predominant in Ivory Coast. The ancestors of the HIV-2 group A and B viruses have been identified in wild sooty mangabey populations in the Tai forest in Ivory Coast, close to the border with Liberia.[93] The other HIV-2 groups have been documented in one or few individuals only. Except for group G and H, group C, D, E, F, and I were isolated in rural areas in Sierra Leone, Liberia, and Ivory Coast, and these viruses are in general more closely related to the SIVsmm strains obtained from sooty mangabeys found in the same area than to any other HIV-2 strain. Molecular clock analysis traced the origin of the epidemic HIV-2 groups A and B to be around 1932 (1906–55) and 1935 (1907–61), respectively.[87] These dates seem to coincide with a political unstable period in Guinea-Bissau, and it has been suggested that civil wars at that time amplified the rapid spread of HIV-2 into the human population.[94]

5. Ongoing Exposure of Humans to a Large Diversity of Simian Immunodeficiency Viruses: Risk for a Novel HIV?

5.1 Exposure to Simian Immunodeficiency Virus–Infected Nonhuman Primates is Still Ongoing

Bushmeat hunting, as a source of animal proteins and income for the family, is a long-standing common component of rural households in the Congo Basin, and more generally throughout sub-Saharan Africa. However, the bushmeat trade has increased significantly during the past decades, due to the expanding logging industry and the

Figure 25.4 HIV-2 is derived from SIVs circulating in sooty mangabeys from West Central Africa. Evolutionary relationship of HIV-2 groups A to I (*black*) and SIVsmm. SIVsmm strains obtained from mangabeys in different regions are colored: *green* (*gray in print versions*) for Ivory Coast, *blue* (*dark gray in print versions*) for Sierra Leone, and *red* (*light gray in print versions*) for Liberia. Letters on the map indicate HIV-2 groups, HIV-2 A and B are epidemic in the western and eastern part of West Africa respectively, the other groups have been identified in single individuals and are colored according to the country of origin. The phylogenetic tree was derived by neighbor-joining analysis of partial (692 bp) *gag* nucleotide sequences from HIV-2 and SIVsmm sequences from West Africa. *Asterisks* represent nonparametric boostrap branch support ≥70%. Horizontal branch lengths are drawn to scale. The geographic range of sooty mangabeys (*Cercocebus atys*) is highlighted in *green*.

Downloaded From Vet-eBooks.com

increasing demand of bushmeat in cities.[95] Indeed, a 2009 study showed that in northern Congo, the bushmeat trade increased together with the increasing presence of logging concessions in the same area.[96] Commercial logging has led to road constructions penetrating remote forest areas, to human migration, and the development of social and economic networks (including those of sex workers), which support this industry. Furthermore, villages around logging concessions have grown from a few hundred to several thousand inhabitants in just a few years, and the number of people entering previously inaccessible forest areas increased significantly over the last decades. Importantly, the HIV prevalences in these rural settings are also increasing, for example, around 20% of women aged between 15 and 35 years are HIV infected in rural forest areas in southern Cameroon.[97,98] A high number of individuals with immune deficiency are thus potentially exposed to new viruses, and recombination between newly introduced SIVs and circulating HIVs can pose an additional risk for the outbreak of a novel HIV epidemic.[97–99]

Studies on primate bushmeat in West and Central Africa showed that bushmeat hunting is not limited to chimpanzees and mangabeys, but the majority of hunted primates are represented by multiple *Cercopithecus* species, colobus monkeys, mandrills, drills, and so on. Moreover, these data revealed ongoing exposure of humans to a plethora of different SIVs.[26,27,100–102] Cross-species transmissions with SIV strains from other primates to humans should thus be considered. It is also to be noted that apes are not only hunted for bushmeat but also for medicinal uses.[103] The socioeconomic changes, which go together with the presence of logging or other industries in remote areas, combined with the SIV prevalence and genetic complexity in wild living primates, suggest that the magnitude of human exposure to SIV has increased, as have the social and environmental conditions that support the emergence and spread of new zoonotic infections.

5.2 Simian Immunodeficiency Virus Prevalences and Cross-Species Transmissions

The chances for cross-species transmissions most likely increase when the frequency of exposure and SIV prevalences are high. A 2009 study in pet monkeys in Cameroon revealed that SIV prevalence in mandrill pets could reach 23%.[104] We have shown among more than 2500 samples in Cameroon and by using SIV lineage-specific Elisas that about 3% of primate bushmeat is SIV infected, but prevalences can vary from 0% to 10% depending on geographic localities, or from 0% to 40% according to species.[26] Interestingly, the lowest prevalences (0–1%) were observed among the most frequently hunted species, thus reducing the risk for cross-species transmissions. However, this situation can be different in other geographic areas, for example, in the DRC a pilot study showed that 20% of the primate bushmeat is infected with the highest prevalences among the most frequently hunted monkeys.[27] Our studies in Cameroon also showed that SIVcpz and SIVgor prevalences in chimpanzee communities infected with the ancestors of HIV-1 M and N and gorilla communities infected with the ancestor of HIV-1 P are among the highest, that is, around 30%.[27,50,70]

Downloaded From Vet-eBooks.com

Moreover, in West Africa, the SIVsmm prevalences of wild mangabeys are around 50%, and at least nine cross-species from mangabeys to humans occurred. In this same region, 50−80% of western red colobus monkeys are also infected with SIVs and are, together with mangabeys, highly represented among primate bushmeat, but no SIVwrc cross-species transmission to humans has been documented yet.[105,106] Western red colobus monkeys also represent 80% of animal proteins among chimpanzees from the *P. t. verus* subspecies in the Tai forest in Ivory Coast, and again no SIVwrc infection could be identified in this chimpanzee subspecies.[107]

5.3 Host Restriction Factors

Several retroviral restriction factors have been identified in humans: APOBEC3G induces lethal hypermutations in the retroviral genome; Trim5alpha proteins restrict the incoming viral capsid; tetherin inhibits the release of viral particles; and SAMHD1 is an antiretroviral protein expressed in the cells of the myeloid lineage that inhibits an early step of the viral life cycle.[108,109] A study reported in 2015 showed the role of the APOBEC3G host restriction factor to explain the absence of SIVwrc from red colobus monkeys in chimpanzees, as mentioned earlier.[110] Similarly, SIVgor resulted from a single introduction of SIVcpz from sympatric chimpanzees, and functional analyses identified APOBEC3G as a barrier for virus transmission from chimpanzees to gorillas.[50]

5.4 Viral Adaptation

The majority of the viruses infecting wild animals are not able to cross the species barrier, adapt to a new host, and spread into the new species. The viral and molecular characteristics that allowed the ancestors of HIV-1 and HIV-2 to cross and adapt to humans are not yet completely identified. More studies are needed to find out why, for example, SIVs from mangabeys, chimpanzees, and gorillas have been transmitted on multiple occasions and not those of other monkeys. Moreover, cross-species transmissions are not always followed by efficient spread into the new populations, as illustrated by HIV-1 versus HIV-2 and among the different HIV-1 groups, and thus depend not only on the virus but also on the host and the environment.[111,112]

Actually, only limited studies showed some viral adaptation, for example, at the Gag-30 position in the p17 region of the *gag* gene, methionine or leucine is present among SIVcpz/SIVgor, but in humans all HIV-1 strains harbor an arginine at this position suggesting an adaptation of the virus to its new host.[113] HIV-1 groups M and O have undergone independent adaptations to acquire resistance to the potent restriction factor tetherin; in group M viruses, Vpu adapted to acquire anti-tetherin activity and in group O, the Nef protein evolved to use a different target within tetherin.[57,114] Nef allows viral persistence in the host but also controls for superactivation of the immune system. However, this latter function is lost in certain SIV lineages and more precisely in the ancestors of the HIV-1/SIVcpz lineage, which could thus have resulted in higher pathogenicity in humans, in contrast to SIVsmm/HIV-2 where this adaptation is not observed and which are less

Downloaded From Vet-eBooks.com

pathogenic.[115] It has been shown in 2013 that passage of SIVs from monkeys through chimpanzees facilitated the subsequent adaptation of HIV-1 to humans, a series of gene loss and adaptation events that generated the chimpanzee precursor of HIV-1, and lowered the species barrier to human infection.[116]

Another study showed a lower viral fitness for HIV-2 compared to HIV-1, and for HIV-1 O versus M, which could also partially explain the lower prevalence and limited spread of HIV-2 and HIV-1 O.[117] Despite progress in our knowledge on viral and host factors, more studies are needed to understand the global picture.

5.5 Human Factors

Human factors also play a major role in the epidemic spread of a new virus, especially among viruses that are transmitted by blood or sexual contacts, as it is the case for HIV. The difference in the localizations of the origin of HIV-1 M (South Cameroon) and the origin of the epidemic (DRC, for HIV-1 M) illustrate the importance of human and environmental factors in the epidemic spread. The main factors involved in the human-to-human spread are sexual risk behavior, high prevalences of sexually transmitted infections, absence of circumcision, and transmission through unsafe blood transfusions and nonsterile needles. These factors associated with sociodemographic factors, such as human density in forest areas, increasing transport between urban and rural areas, human migration, urbanization, and increase in commercial sex, were in favor of an epidemic spread of the virus.

The early spread of HIV-1 in the human population has been reconstructed in 2014 with statistical phylodynamic models on large collections of sequence data from Central Africa.[86] The study confirmed that the HIV-1 group M pandemic started in Kinshasa around the mid-1920s. The ancestor of HIV-1 M arrived most likely from southern Cameroon in Kinshasa via the Sangha and Congo rivers during the early colonial period.[86,118] The subsequent spatial expansion in Central Africa is associated with transportation networks, railway, and fluvial. Iatrogenic interventions and/or changes in sexual behavior were critical for the pandemic growth of group M. A unique combination of circumstances during a particular spatial and sociohistorical window thus allowed the establishment, spatial expansion, and epidemic growth of HIV-1 group M to pandemic proportions. The fact that group O viruses have not spread even more widely in the human population could thus reflect the absence of epidemiological opportunity during the early stages of the pandemic expansion of AIDS 50 years ago.[86]

5.6 Ongoing Cross-Species Transmissions From Other Retroviruses From Primates to Humans

Simian foamy virus (SFV) is infecting primates at high levels: 70% of primates in captivity, 97% of wild western red colobus monkeys, and between 44% and 100% of wild chimpanzees.[119–121] SFVs have been documented in primate care workers in the United States, in hunters in Cameroon and Gabon, and in women living in

Downloaded From Vet-eBooks.com

rural DRC.[122−125] Moreover, SFV prevalence can reach 25% in persons who reported ape bites.[126] Epidemiological studies as reported in 2013 and 2015 in New World primates (NWPs) also showed a wide distribution of distinct SFVs, highlighting the risk of potential zoonotic transmission in this area, which is still poorly studied.[127,128] However, human-to-human transmission or an SFV epidemic has not been documented yet, and no pathogenicity associated with SFV infection has been observed in humans. In this example, cross-species transmissions were thus possible, but the viral adaptation was insufficient to allow a spread of the virus into the new host.

Exposure to NHP also allowe the emergence of four types of human T-lymphotropic virus (HTLV) type 1 to 4 in humans, all of them with the simian counterparts (STLV-1−4) already identified.[129−131] Furthermore, several studies among hunters in Central Africa showed ongoing cross-species transmissions of STLVs.[132−134] These T-lymphotropic viruses are thus able to cross the species barrier, and certain variants were able to spread into the human population leading to HTLV-1 and HTLV-2, which are endemic in certain parts of the world.

The ongoing transmissions of SFV and STLV highlight the risk for potential emergence of a new SIV into the human population. The discoveries in 2009 of a new HIV-1, group P, and in 2013 of a new HIV-2, group I, are additional demonstrations of ongoing zoonotic transmissions.[91,135,136]

6. Conclusion

Today, we have a clear picture of the origin of HIV and the seeds of the AIDS pandemic. The current HIV-1 epidemic provides evidence for the extraordinary impact that can result from a single primate lentiviral zoonotic transmission event. Despite the fact that the first AIDS cases have been observed around 1980 in the United States, the virus circulated already early in the 20th century in the human population in Central Africa (Fig. 25.1). Currently, at least, 13 SIV cross-species transmissions occurred, 9 for HIV-2 and 4 for HIV-1, but most likely several others occurred in the past, which remained unrecognized since some viruses were not able to adapt to the new host or since the environment was not suitable for epidemic spread. Because humans are still exposed to a plethora of primate lentiviruses through hunting and handling of primate bushmeat, the possibility of additional cross-species transfers of primate lentiviruses from other primate species in addition to chimpanzees, gorillas, or sooty mangabeys has to be considered. One major public health implication is that these SIVs strains are not always recognized by commercial HIV-1/HIV-2 screening assays. As a consequence, due to the long incubation period, human infection with such variants can go unrecognized for several years and lead to another epidemic. In addition, social and environmental factors associated with epidemic spread are now present in rural forest areas. Identification of SIVs in wild primates can serve as sentinels by signaling which pathogens may be a risk for humans and allow the development of serological and molecular assays to detect transmissions with other SIVs in humans.

Downloaded From Vet-eBooks.com

References

1. Centers for Disease Control (CDC). Pneumocystis pneumonia − Los Angeles. *MMWR Morb Mortal Wkly Rep* 1981;**30**(21):250−2.
2. Barré-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 1983;**220**:868−71.
3. Ellrodt A, Barre-Sinoussi F, Le Bras P, Nugeyre M, Palazzo L, Rey F, et al. Isolation of human T-lymphotropic retrovirus (LAV) from Zairian married couple, one with AIDS, one with prodromes. *Lancet* 1984;**1**:1383−5.
4. Piot P, Quinn TC, Taelman H, Feinsod FM, Minlangu KB, Wobin O, et al. Acquired immunodeficiency syndrome in a heterosexual population in Zaire. *Lancet* 1984;**2**:65−9.
5. Van de Perre P, Rouvroy D, Lepage P, Bogaerts J, Kestelyn P, Kayihigi J, et al. Acquired immunodeficiency syndrome in Rwanda. *Lancet* 1984;**2**:62−5.
6. Buvé A, Bishikwabo-Nsarhaza K, Mutangadura G. The spread and effect of HIV-1 infection in sub-Saharan Africa. *Lancet* 2002;**359**:2011−7.
7. Vidal N, Peeters M, Mulanga-Kabeya C, Nzilambi N, Robertson D, Ilunga W, et al. Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa. *J Virol* 2000;**74**:10498−507.
8. Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, Bunce M, et al. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature* 2008;**455**:661−4.
9. Gilbert M, Rambaut A, Wlasiuk G, Spira TJ, Pitchenik AE, Worobey M. The emergence of HIV/AIDS in the Americas and beyond. *Proc Natl Acad Sci USA* 2007;**104**:18566−70.
10. Low-Beer D. The distribution of early acquired immune deficiency syndrome cases and conditions for the establishment of new epidemics. *Philos Trans R Soc Lond B Biol Sci* 2001;**356**:927−31.
11. UNAIDS/WHO. *The gap report, UNAIDS/JC2656. United Nations programme on HIV/AIDS (UNAIDS) and World Health Organization (WHO).* 2014.
12. Barin F, M'Boup S, Denis F, Kanki P, Allan JS, Lee TH, et al. Serological evidence for virus related to simian T-lymphotropic retrovirus III in residents of west Africa. *Lancet* 1985;**2**:1387−9.
13. Clavel F, Guétard D, Brun-Vézinet F, Chamaret S, Rey MA, Santos-Ferreira MO, et al. Isolation of a new human retrovirus from West African patients with AIDS. *Science* 1986; **233**:343−6.
14. Kanki PJ, Barin F, M'Boup S, Allan JS, Romet-Lemonne JL, Marlink R. New human T-lymphotropic retrovirus related to simian T-lymphotropic virus type III (STLV-IIIAGM). *Science* 1986;**232**:238−43.
15. Daniel MD, Letvin NL, King NW, Kannagi M, Sehgal PK, Hunt RD, et al. Isolation of T-cell tropic HTLV-III-like retrovirus from macaques. *Science* 1985;**228**:1201−4.
16. Henrickson RV, Maul DH, Osborn KG, Sever JL, Madden DL, Ellingsworth LR, et al. Epidemic of acquired immunodeficiency in rhesus monkeys. *Lancet* 1983;**1**: 388−90.
17. Mansfield KG, Lerch NW, Gardner MB, Lackner AA. Origins of simian immunodeficiency virus infection in macaques at the New England regional primate Research center. *J Med Primatol* 1995;**24**:116−22.
18. Lowenstine LJ, Pedersen NC, Higgins J, Pallis KC, Uyeda A, Marx P, et al. Seroepidemiologic survey of captive Old-World primates for antibodies to human and simian

Downloaded From Vet-eBooks.com

retroviruses, and isolation of a lentivirus from sooty mangabeys (*Cercocebus atys*). *Int J Cancer* 1986;**38**:563−74.
19. Apetrei C, Kaur A, Lerche NW, Metzger M, Pandrea I, Hardcastle J, et al. Molecular epidemiology of simian immunodeficiency virus SIVsm in U.S. primate centers unravels the origin of SIVmac and SIVstm. *J Virol* 2005;**79**:8991−9005.
20. Murphey-Corb M, Martin LN, Rangan S, Baskin GB, Gormus BJ, Wolf RH, et al. Isolation of an HTLV-III-related retrovirus from macaques with simian AIDS and its possible origin in asymptomatic mangabeys. *Nature* 1986;**321**:435−7.
21. Gao F, Yue L, White AT, Pappas PG, Barchue J, Hanson AP, et al. Human infection by genetically diverse SIVSM-related HIV-2 in west Africa. *Nature* 1992; **358**:495−9.
22. Hirsch VM, Olmsted RA, Murphey-Corb M, Purcell RH, Johnson PR. An African primate lentivirus (SIVsm) closely related to HIV-2. *Nature* 1989;**339**:389−92.
23. Peeters M, D'Arc M, Delaporte E. Origin and diversity of human retroviruses. *AIDS Rev* 2014;**16**:23−34.
24. Simon F, Souquière S, Damond F, Kfutwah A, Makuwa M, Leroy E, et al. Synthetic peptide strategy for the detection of and discrimination among highly divergent primate lentiviruses. *AIDS Res Hum Retroviruses* 2001;**17**:937−52.
25. Aghokeng AF, Liu W, Bibollet-Ruche F, Loul S, Mpoudi-Ngole E, Laurent C, et al. Widely varying SIV prevalence rates in naturally infected primate species from Cameroon. *Virology* 2006;**345**:174−89.
26. Aghokeng AF, Ayouba A, Mpoudi-Ngole E, Loul S, Liegeois F, Delaporte E, et al. Extensive survey on the prevalence and genetic diversity of SIVs in primate bushmeat provides insights into risks for potential new cross-species transmissions. *Infect Genet Evol* 2010;**10**:386−96.
27. Ahuka-Mundeke S, Ayouba A, Mbala-Kingebeni P, Liegeois F, Esteban A, Lunguya-Metila O, et al. Novel multiplexed HIV/simian immunodeficiency virus antibody detection assay. *Emerg Infect Dis* 2011;**17**:2277−86.
28. Ayouba A, Duval L, Liégeois F, Ngin S, Ahuka-Mundeke S, Switzer WM, et al. Nonhuman primate retroviruses from Cambodia: high simian foamy virus prevalence, identification of divergent STLV-1 strains and no evidence of SIV infection. *Infect Genet Evol* 2013;**18**:325−34.
29. Locatelli S, Peeters M. Cross-species transmission of simian retroviruses: how and why they could lead to the emergence of new diseases in the human population. *AIDS* 2012;**26**: 659−73.
30. Worobey M, Telfer P, Souquière S, Hunter M, Coleman C, Metzger M, et al. Island biogeography reveals the deep history of SIV. *Science* 2010;**329**:1487.
31. Ma D, Jasinska A, Kristoff J, Grobler JP, Turner T, Jung Y, et al. SIVagm infection in wild African green monkeys from South Africa: epidemiology, natural history, and evolutionary considerations. *PLoS Pathog* 2013;**9**:e1003011.
32. Silvestri G, Paiardini M, Pandrea I, Lederman MM, Sodora DL. Understanding the benign nature of SIV infection in natural hosts. *J Clin Invest* 2007;**117**:3148−54.
33. Keele BF, Jones JH, Terio KA, Estes JD, Rudicell RS, Wilson ML, et al. Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz. *Nature* 2009;**460**:515−9.
34. Liovat AS, Jacquelin B, Ploquin MJ, Barré-Sinoussi F, Müller-Trutwin MC. African non human primates infected by SIV − why don't they get sick? Lessons from studies on the early phase of non-pathogenic SIV infection. *Curr HIV Res* 2009;**7**:39−50.

Downloaded From Vet-eBooks.com

35. Silvestri G, Sodora DL, Koup RA, Paiardini M, O'Neil SP, McClure HM, et al. Nonpathogenic SIV infection of sooty mangabeys is characterized by limited bystander immunopathology despite chronic high-level viremia. *Immunity* 2003;**18**:441−52.
36. Ling B, Apetrei C, Pandrea I, Veazey RS, Lackner AA, Gormus B, et al. Classic AIDS in a sooty mangabey after an 18-year natural infection. *J Virol* 2004;**78**:8902−8.
37. Etienne L, Nerrienet E, LeBreton M, Bibila GT, Foupouapouognigni Y, Rousset D, et al. Characterization of a new simian immunodeficiency virus strain in a naturally infected *Pan troglodytes troglodytes* chimpanzee with AIDS related symptoms. *Retrovirology* 2011; **13**(8):4.
38. Rudicell RS, Holland Jones J, Wroblewski EE, Learn GH, Li Y, Robertson JD, et al. Impact of simian immunodeficiency virus infection on chimpanzee population dynamics. *PLoS Pathog* 2010;**6**:e1001116.
39. Wertheim JO, Worobey M. A challenge to the ancient origin of SIVagm based on African green monkey mitochondrial genomes. *PLoS Pathog* 2007;**3**:e95.
40. Beer BE, Bailes E, Goeken R, Dapolito G, Coulibaly C, Norley SG, et al. Simian immunodeficiency virus (SIV) from sun-tailed monkeys (*Cercopithecus solatus*): evidence for host-dependent evolution of SIV within the *C. lhoesti* superspecies. *J Virol* 1999;**73**: 7734−44.
41. Bibollet-Ruche F, Bailes E, Gao F, Pourrut X, Barlow KL, Clewley JP, et al. New simian immunodeficiency virus infecting De Brazza's monkeys (*Cercopithecus neglectus*): evidence for a cercopithecus monkey virus clade. *J Virol* 2004;**78**:7748−62.
42. Ayouba A, Njouom R, Chia JE, Ahuka-Mundeke S, Kfutwah A, Ngole EM, et al. Molecular characterization of a new mosaic Simian Immunodeficiency Virus in a naturally infected tantalus monkey (*Chlorocebus tantalus*) from Cameroon: a challenge to the virus-host co-evolution of SIVagm in African green monkeys. *Infect Genet Evol* 2015;**30**: 65−73.
43. Charleston MA, Robertson DL. Preferential host switching by primate lentiviruses can account for phylogenetic similarity with the primate phylogeny. *Syst Biol* 2002;**51**: 528−35.
44. Bibollet-Ruche F, Galat-Luong A, Cuny G, Sarni-Manchado P, Galat G, Durand JP, et al. Simian immunodeficiency virus infection in a patas monkey (*Erythrocebus patas*): evidence for cross-species transmission from African green monkeys (*Cercopithecus aethiops sabaeus*) in the wild. *J Gen Virol* 1996;**77**:773−81.
45. Jin MJ, Rogers J, Phillips-Conroy JE, Allan JS, Desrosiers RC, Shaw GM, et al. Infection of a yellow baboon with simian immunodeficiency virus from African green monkeys: evidence for cross-species transmission in the wild. *J Virol* 1994;**68**:8454−60.
46. Van Rensburg EJ, Engelbrecht S, Mwenda J, Laten JD, Robson BA, Stander T, et al. Simian immunodeficiency viruses (SIVs) from eastern and southern Africa: detection of a SIVagm variant from a chacma baboon. *J Gen Virol* 1998;**79**:1809−14.
47. Aghokeng AF, Bailes E, Loul S, Courgnaud V, Mpoudi-Ngolle E, Sharp PM, et al. Full-length sequence analysis of SIVmus in wild populations of mustached monkeys (*Cercopithecus cephus*) from Cameroon provides evidence for two co-circulating SIVmus lineages. *Virology* 2007;**360**:407−18.
48. Liégeois F, Schmidt F, Boué V, Butel C, Mouacha F, Ngari P, et al. Full-length genome analyses of two new simian immunodeficiency virus (SIV) strains from mustached monkeys (*C. cephus*) in Gabon illustrate a complex evolutionary history among the SIVmus/mon/gsn lineage. *Viruses* 2014;**6**:2880−98.
49. Bailes E, Gao F, Bibollet-Ruche F, Courgnaud V, Peeters M, Marx PA, et al. Hybrid origin of SIV in chimpanzees. *Science* 2003;**300**:1713.

Downloaded From Vet-eBooks.com

50. D'arc M, Ayouba A, Esteban A, Learn GH, Boué V, Liegeois F, et al. Origin of the HIV-1 group O epidemic in western lowland gorillas. *Proc Natl Acad Sci USA* 2015;**112**: E1343−52.
51. Takehisa J, Kraus MH, Ayouba A, Bailes E, Van Heuverswyn F, Decker JM, et al. Origin and biology of simian immunodeficiency virus in wild-living western gorillas. *J Virol* 2009;**83**:1635−48.
52. Van Heuverswyn F, Li Y, Neel C, Bailes E, Keele BF, Liu W, et al. Human immunodeficiency viruses: SIV infection in wild gorillas. *Nature* 2006;**444**:164.
53. Souquière S, Bibollet-Ruche F, Robertson DL, Makuwa M, Apetrei C, Onanga R, et al. Wild *Mandrillus sphinx* are carriers of two types of lentivirus. *J Virol* 2001;**75**:7086−96.
54. Hahn BH, Shaw GM, De Cock KM, Sharp PM. AIDS as a zoonosis: scientific and public health implications. *Science* 2000;**287**:607−14.
55. De Leys R, Vanderborght B, Vanden Haesevelde M, Heyndrickx L, van Geel A, Wauters C, et al. Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of west-central African origin. *J Virol* 1990;**64**: 1207−16.
56. Ayouba A, Mauclère P, Martin PM, Cunin P, Mfoupouendoun J, Njinku B, et al. HIV-1 group O infection in Cameroon, 1986 to 1998. *Emerg Infect Dis* 2001;**7**: 466−7.
57. Kluge SF, Mack K, Iyer SS, Pujol FM, Heigele A, Learn GH, et al. Nef proteins of epidemic HIV-1 group O strains antagonize human tetherin. *Cell Host Microbe* 2014;**16**: 639−50.
58. Villabona-Arenas CJ, Domyeum J, Mouacha F, Butel C, Delaporte E, Peeters M, et al. HIV-1 group O infection in Cameroon from 2006 to 2013: prevalence, genetic diversity, evolution and public health challenges. *Infect Genet Evol* 2015;**36**:210−6.
59. Delaugerre C, De Oliveira F, Lascoux-Combe C, Plantier JC, Simon F. HIV-1 group N: travelling beyond Cameroon. *Lancet* 2011;**378**:1894.
60. Mourez T, Simon F, Plantier JC. Non-M variants of human immunodeficiency virus type 1. *Clin Microbiol Rev* 2013;**26**:448−61.
61. Peeters M, Honoré C, Huet T, Bedjabaga L, Ossari S, Bussi P, et al. Isolation and partial characterization of an HIV-related virus occurring naturally in chimpanzees in Gabon. *AIDS* 1989;**3**:625−30.
62. Huet T, Cheynier R, Meyerhans A, Roelants G, Wain-Hobson S. Genetic organization of a chimpanzee lentivirus related to HIV-1. *Nature* 1990;**345**:356−9.
63. Peeters M, Fransen K, Delaporte E, Van den Haesevelde M, Gershy-Damet GM, Kestens L, et al. Isolation and characterization of a new chimpanzee lentivirus (simian immunodeficiency virus isolate cpz-ant) from a wild-captured chimpanzee. *AIDS* 1992;**6**: 447−51.
64. Vanden Haesevelde MM, Peeters M, Jannes G, Janssens W, van der Groen G, Sharp PM, et al. Sequence analysis of a highly divergent HIV-1-related lentivirus isolated from a wild captured chimpanzee. *Virology* 1996;**221**:346−50.
65. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, et al. Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature* 1999;**397**: 436−41.
66. Santiago ML, Rodenburg CM, Kamenya S, Bibollet-Ruche F, Gao F, Bailes E, et al. SIVcpz in wild chimpanzees. *Science* 2002;**295**:465.
67. Santiago ML, Lukasik M, Kamenya S, Li Y, Bibollet-Ruche F, Bailes E, et al. Foci of endemic simian immunodeficiency virus infection in wild-living eastern chimpanzees (*Pan troglodytes schweinfurthii*). *J Virol* 2003;**77**:7545−62.

Downloaded From Vet-eBooks.com

68. Santiago ML, Bibollet-Ruche F, Bailes E, Kamenya S, Muller MN, Lukasik M, et al. Amplification of a complete simian immunodeficiency virus genome from fecal RNA of a wild chimpanzee. *J Virol* 2003;**77**:2233–42.
69. Boué V, Locatelli S, Boucher F, Ayouba A, Butel C, Esteban A, et al. High rate of Simian Immunodeficiency Virus (SIV) infections in wild chimpanzees in northeastern Gabon. *Viruses* 2015;**7**:4997–5015.
70. Keele BF, Van Heuverswyn F, Li Y, Bailes E, Takehisa J, Santiago ML, et al. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science* 2006;**313**:523–6.
71. Li Y, Ndjango JB, Learn GH, Ramirez MA, Keele BF, Bibollet-Ruche F, et al. Eastern chimpanzees, but not bonobos, represent a simian immunodeficiency virus reservoir. *J Virol* 2012;**86**:10776–91.
72. Van Heuverswyn F, Li Y, Bailes E, Neel C, Lafay B, Keele BF, et al. Genetic diversity and phylogeographic clustering of SIVcpz*Ptt* in wild chimpanzees in Cameroon. *Virology* 2007;**368**:155–71.
73. Worobey M, Santiago ML, Keele BF, Ndjango J, Joy JB, Labama BL, et al. Origin of AIDS: contaminated polio vaccine theory refuted. *Nature* 2004;**428**:820.
74. Prince AM, Brotman B, Lee DH, Andrus L, Valinsky J, Marx P. Lack of evidence for HIV type 1-related SIVcpz infection in captive and wild chimpanzees (*Pan troglodytes verus*) in West Africa. *AIDS Res Hum Retroviruses* 2002;**18**:657–60.
75. Switzer WM, Parekh B, Shanmugam V, Bhullar V, Phillips S, Ely JJ, et al. The epidemiology of simian immunodeficiency virus infection in a large number of wild- and captive-born chimpanzees: evidence for a recent introduction following chimpanzee divergence. *AIDS Res Hum Retroviruses* 2005;**21**:335–42.
76. Head JS, Boesch C, Makaga L, Robbins MM. Sympatric chimpanzees (*Pan troglodytes troglodytes*) and gorillas (*Gorilla gorilla gorilla*) in Loango National Park, Gabon: dietary composition, seasonality, and intersite comparisons. *Int J Primatol* 2011;**32**:755–75.
77. Morgan D, Sanz C. Chimpanzee feeding ecology and comparisons with sympatric gorillas in the Goualougo Triangle, Republic of Congo. In: *Feeding ecology in apes and other primates*. Cambridge, UK: Cambridge University Press; 2006. p. 97–122.
78. Stanford CB, Nkurunungi JB. Behavioral ecology of sympatric chimpanzees and gorillas in Bwindi Impenetrable National Park, Uganda: Diet. *Int J Primatol* 2003;**24**:901–18.
79. Lyons S, Sharp C, LeBreton M, Djoko CF, Kiyang JA, Lankester F, et al. Species association of hepatitis B virus (HBV) in non-human apes; evidence for recombination between gorilla and chimpanzee variants. *PLoS One* 2012;**7**:e33430.
80. Walsh PD, Breuer T, Sanz C, Morgan D, Doran-Sheehy D. Potential for Ebola transmission between gorilla and chimpanzee social groups. *Am Nat* 2007;**169**:684–9.
81. Neel C, Etienne L, Li Y, Takehisa J, Rudicell RS, Bass IN, et al. Molecular epidemiology of simian immunodeficiency virus infection in wild-living gorillas. *J Virol* 2010;**84**: 1464–76.
82. Brennan CA, Bodelle P, Coffey R, Devare SG, Golden A, Hackett J, et al. The prevalence of diverse HIV-1 strains was stable in Cameroonian blood donors from 1996 to 2004. *J Acquir Immune Defic Syndr* 2008;**49**:432–9.
83. Vessière A, Rousset D, Kfutwah A, Leoz M, Depatureaux A, Simon F, et al. Diagnosis and monitoring of HIV-1 group O-infected patients in Cameroun. *J Acquir Immune Defic Syndr* 2010;**53**:107–10.
84. Zhu T, Korber BT, Nahmias AJ, Hooper E, Sharp PM, Ho DD. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* 1998;**391**: 594–7.

Downloaded From Vet-eBooks.com

85. Jonassen TO, Stene-Johansen K, Berg ES, Hungnes O, Lindboe CF, Frøland SS, et al. Sequence analysis of HIV-1 group O from Norwegian patients infected in the 1960s. *Virology* 1997;**231**:43−7.
86. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, et al. HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science* 2014;**346**:56−61.
87. Wertheim JO, Worobey M. Dating the age of the SIV lineages that gave rise to HIV-1 and HIV-2. *PLoS Comput Biol* 2009;**5**:e1000377.
88. Van Tienen C, van der Loeff M, Zaman S, Vincent T, Sarge-Njie R, Peterson I, et al. Two distinct epidemics: the rise of HIV-1 and decline of HIV-2 infection between 1990 and 2007 in rural Guinea-Bissau. *J Acquir Immune Defic Syndr* 2010;**53**:640−7.
89. Gottlieb GS, Hawes SE, Agne HD, Stern JE, Critchlow CW, Kiviat NB, et al. Lower levels of HIV RNA in semen in HIV-2 compared with HIV-1 infection: implications for differences in transmission. *AIDS* 2006;**20**:895−900.
90. Hawes SE, Sow PS, Stern JE, Critchlow CW, Gottlieb GS, Kiviat NB. Lower levels of HIV-2 than HIV-1 in the female genital tract: correlates and longitudinal assessment of viral shedding. *AIDS* 2008;**22**:2517−25.
91. Ayouba A, Akoua-Koffi C, Calvignac-Spencer S, Esteban A, Locatelli S, Li H, et al. Evidence for continuing cross-species transmission of SIVsmm to humans: characterization of a new HIV-2 lineage in rural Côte d'Ivoire. *AIDS* 2013;**27**:2488−91.
92. Damond F, Worobey M, Campa P, Farfara I, Colin G, Matheron S, et al. Identification of a highly divergent HIV type 2 and proposal for a change in HIV type 2 classification. *AIDS Res Hum Retroviruses* 2004;**20**:666−72.
93. Santiago ML, Range F, Keele BF, Li Y, Bailes E, Bibollet-Ruche F, et al. Simian immunodeficiency virus infection in free-ranging sooty mangabeys (*Cercocebus atys atys*) from the Taï Forest, Côte d'Ivoire: implications for the origin of epidemic human immunodeficiency virus type 2. *J Virol* 2005;**79**:12515−27.
94. Lemey P, Pybus OG, Wang B, Saksena NK, Salemi M, Vandamme AM. Tracing the origin and history of the HIV-2 epidemic. *Proc Natl Acad Sci USA* 2003;**100**:6588−92.
95. Laporte NT, Stabach JA, Grosch R, Lin TS, Goetz SJ. Expansion of industrial logging in central Africa. *Science* 2007;**316**:1451.
96. Poulsen JR, Clark CJ, Mavah G, Elkan PW. Bushmeat supply and consumption in a tropical logging concession in northern Congo. *Conserv Biol* 2009;**23**:1597−608.
97. Chia JE, Aghokeng A, Guichet E, Ayouba A, Ahuka-Mundeke S, Vidal N, et al. Ongoing cross-species transmission of simian retroviruses and high HIV prevalence in Cameroon. In: *CROI 2014-Program and Abstracts. Presented at the Conference on retroviruses and opportunistic infections 2014*. USA MA: Boston; 2014. p. 222.
98. Laurent C, Bourgeois A, Mpoudi M, Butel C, Peeters M, Mpoudi-Ngolé E, et al. Commercial logging and HIV epidemic, rural Equatorial Africa. *Emerg Infect Dis* 2004;**10**: 1953−6.
99. LeBreton M, Yang O, Tamoufe U, Mpoudi-Ngole E, Torimiro JN, Djoko CF, et al. Exposure to wild primates among HIV-infected persons. *Emerg Infect Dis* 2007;**13**: 1579−82.
100. Liégeois F, Boué V, Mouacha F, Butel C, Ondo BM, Pourrut X, et al. New STLV-3 strains and a divergent SIVmus strain identified in non-human primate bushmeat in Gabon. *Retrovirology* 2012;**9**:28.
101. Mossoun A, Pauly M, Akoua-Koffi C, Couacy-Hymann E, Leendertz SAJ, et al. Contact to non-human primates and risk factors for zoonotic disease emergence in the Taï region, Côte d'ivoire. *Ecohealth* 2015;**12**:580−91.

Downloaded From Vet-eBooks.com

102. Peeters M, Courgnaud V, Abela B, Auzel P, Pourrut X, Bibollet-Ruche F, et al. Risk to human health from a plethora of simian immunodeficiency viruses in primate bushmeat. *Emerg Infect Dis* 2002;**8**:451–7.
103. Meder A. Gorillas in African culture and medicine. *Gorilla J* 1999;**18**:11–5.
104. Ndembi N, Kaptue L, Ido E. Exposure to SIVmnd-2 in southern Cameroon: public health implications. *AIDS Rev* 2009;**11**:135–9.
105. Leendertz SAJ, Junglen S, Hedemann C, Goffe A, Calvignac S, Boesch C, et al. High prevalence, coinfection rate, and genetic diversity of retroviruses in wild red colobus monkeys (*Piliocolobus badius badius*) in Tai National Park, Cote d'Ivoire. *J Virol* 2010; **84**:7427–36.
106. Locatelli S, Liegeois F, Lafay B, Roeder AD, Bruford MW, Formenty P, et al. Prevalence and genetic diversity of simian immunodeficiency virus infection in wild-living red colobus monkeys (*Piliocolobus badius badius*) from the Taï forest, Côte d'Ivoire SIVwrc in wild-living western red colobus monkeys. *Infect Genet Evol* 2008;**8**:1–14.
107. Leendertz SAJ, Locatelli S, Boesch C, Kücherer C, Formenty P, Liegeois F, et al. No evidence for transmission of SIVwrc from western red colobus monkeys (*Piliocolobus badius badius*) to wild West African chimpanzees (*Pan troglodytes verus*) despite high exposure through hunting. *BMC Microbiol* 2011;**11**:24.
108. Kirchhoff F. Immune evasion and counteraction of restriction factors by HIV-1 and other primate lentiviruses. *Cell Host Microbe* 2010;**8**:55–67.
109. Laguette N, Sobhian B, Casartelli N, Ringeard M, Chable-Bessia C, Ségéral E, et al. SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature* 2011;**474**:654–7.
110. Etienne L, Bibollet-Ruche F, Sudmant PH, Wu LI, Hahn BH, Emerman M. The role of the antiviral APOBEC3 gene family in protecting chimpanzees against lentiviruses from monkeys. *PLoS Pathog* 2015;**11**:e1005149.
111. Chastel C. Emergential success: a new concept for a better appraisal of viral emergences and reemergences. *Acta Virol* 2000;**44**:375–6.
112. Lloyd-Smith JO, George D, Pepin KM, Pitzer VE, Pulliam JRC, Dobson AP, et al. Epidemic dynamics at the human-animal interface. *Science* 2009;**326**:1362–7.
113. Wain LV, Bailes E, Bibollet-Ruche F, Decker JM, Keele BF, Van Heuverswyn F, et al. Adaptation of HIV-1 to its human host. *Mol Biol Evol* 2007;**24**:1853–60.
114. Sauter D, Schindler M, Specht A, Landford WN, Münch J, Kim KA, et al. Tetherin-driven adaptation of Vpu and Nef function and the evolution of pandemic and nonpandemic HIV-1 strains. *Cell Host Microbe* 2009;**6**:409–21.
115. Schindler M, Münch J, Kutsch O, Li H, Santiago ML, Bibollet-Ruche F, et al. Nef-mediated suppression of T cell activation was lost in a lentiviral lineage that gave rise to HIV-1. *Cell* 2006;**125**:1055–67.
116. Etienne L, Hahn BH, Sharp PM, Matsen FA, Emerman M. Gene loss and adaptation to hominids underlie the ancient origin of HIV-1. *Cell Host Microbe* 2013; **14**:85–92.
117. Ariën KK, Abraha A, Quiñones-Mateu ME, Kestens L, Vanham G, Arts EJ. The replicative fitness of primary human immunodeficiency virus type 1 (HIV-1) group M, HIV-1 group O, and HIV-2 isolates. *J Virol* 2005;**79**:8979–90.
118. Sharp PM, Hahn BH. AIDS: prehistory of HIV-1. *Nature* 2008;**455**:605–6.
119. Hussain AI, Shanmugam V, Bhullar VB, Beer BE, Vallet D, et al. Screening for simian foamy virus infection by using a combined antigen Western blot assay: evidence for a wide distribution among Old World primates and identification of four new divergent viruses. *Virology* 2003;**309**:248–57.

Downloaded From Vet-eBooks.com

120. Goldberg TL, Sintasath DM, Chapman CA, Cameron KM, Karesh WB, Tang S, et al. Coinfection of Ugandan red colobus (*Procolobus [Piliocolobus] rufomitratus tephrosceles*) with novel, divergent delta-, lenti-, and spumaretroviruses. *J Virol* 2009;**83**: 11318−29.
121. Liu W, Worobey M, Li Y, Keele BF, Bibollet-Ruche F, Guo Y, et al. Molecular ecology and natural history of simian foamy virus infection in wild-living chimpanzees. *PLoS Pathog* 2008;**4**:e1000097.
122. Betsem E, Rua R, Tortevoye P, Froment A, Gessain A. Frequent and recent human acquisition of simian foamy viruses through apes' bites in central Africa. *PLoS Pathog* 2011;**7**:e1002306.
123. Mouinga-Ondémé A, Caron M, Nkoghé D, Telfer P, Marx P, Saïb A, et al. Cross-species transmission of simian foamy virus to humans in rural Gabon, Central Africa. *J Virol* 2012; **86**:1255−60.
124. Switzer WM, Tang S, Ahuka-Mundeke S, Shankar A, Hanson DL, Zheng H, et al. Novel simian foamy virus infections from multiple monkey species in women from the Democratic Republic of Congo. *Retrovirology* 2012;**9**:100.
125. Wolfe ND, Switzer W, Carr JK, Bhullar VB, Shanmugam V, Tamoufe U, et al. Naturally acquired simian retrovirus infections in central African hunters. *Lancet* 2004;**363**:932−7.
126. Calattini S, Betsem EBA, Froment A, Mauclère P, Tortevoye P, Schmitt C, et al. Simian foamy virus transmission from apes to humans, rural Cameroon. *Emerg Infect Dis* 2007; **13**:1314−20.
127. Muniz CP, Troncoso LL, Moreira MA, Soares EA, Pissinatti A, Bonvicino CR, et al. Identification and characterization of highly divergent simian foamy viruses in a wide range of new world primates from Brazil. *PLoS One* 2013;**8**:e67568.
128. Muniz CP, Jia H, Shankar A, Troncoso LL, Augusto AM, Farias E, et al. An expanded search for simian foamy viruses (SFV) in Brazilian New World primates identifies novel SFV lineages and host age-related infections. *Retrovirology* 2015;**12**:94.
129. LeBreton M, Switzer WM, Djoko CF, Gillis A, Jia H, Sturgeon MM, et al. A gorilla reservoir for human T-lymphotropic virus type 4. *Emerg Microbes Infect* 2014;**3**:e7.
130. Switzer WM, Salemi M, Qari SH, Jia H, Gray RR, Katzourakis A, et al. Ancient, independent evolution and distinct molecular features of the novel human T-lymphotropic virus type 4. *Retrovirology* 2009;**6**:9.
131. Wolfe ND, Heneine W, Carr JK, Garcia AD, Shanmugam V, Tamoufe U, et al. Emergence of unique primate T-lymphotropic viruses among central African bushmeat hunters. *Proc Natl Acad Sci USA* 2005;**102**:7994−9.
132. Calvignac-Spencer S, Adjogoua EV, Akoua-Koffi C, Hedemann C, Schubert G, Ellerbrok H, et al. Origin of human T-lymphotropic virus type 1 in rural Côte d'Ivoire. *Emerg Infect Dis* 2012;**18**:830−3.
133. Sintasath DM, Wolfe ND, Zheng HQ, LeBreton M, Peeters M, Tamoufe U, et al. Genetic characterization of the complete genome of a highly divergent simian T-lymphotropic virus (STLV) type 3 from a wild *Cercopithecus mona* monkey. *Retrovirology* 2009;**6**:97.
134. Zheng H, Wolfe ND, Sintasath DM, Tamoufe U, Lebreton M, Djoko CF, et al. Emergence of a novel and highly divergent HTLV-3 in a primate hunter in Cameroon. *Virology* 2010; **401**:137−45.
135. Plantier JC, Leoz M, Dickerson JE, De Oliveira F, Cordonnier F, Lemée V, et al. A new human immunodeficiency virus derived from gorillas. *Nat Med* 2009;**15**: 871−2.
136. Vallari A, Holzmayer V, Harris B, Yamaguchi J, Ngansop C, Makamche F, et al. Confirmation of putative HIV-1 group P in Cameroon. *J Virol* 2011;**85**:1403−7.

Downloaded From Vet-eBooks.com

This page intentionally left blank

Downloaded From Vet-eBooks.com

Evolution of SARS Coronavirus and the Relevance of Modern Molecular Epidemiology

26

Z. Shi[1], *L.-F. Wang*[2]
[1]Chinese Academy of Sciences (CAS), Wuhan, China; [2]Duke-NUS Medical School, Singapore, Singapore

1. A Brief History of SARS

As outlined in Table 26.1, the first reported case of "atypical pneumonia," now known as Severe Acute Respiratory Syndrome or SARS, occurred in Guangzhou, Guangdong province, China, on November 16, 2002. Before the end of February 2003, a total of 11 index cases occurred independently in nine cities of Guangdong Province, which forms the early phase of the SARS epidemic.[1] These index cases spread the virus to their close relatives and hospital staffs and provided the early demonstration of the respiratory transmission mode of the disease. The clinical symptoms of SARS are nonspecific. The index cases all began to have fever higher than 38°C and displayed common respiratory symptoms, such as cough, headache, and shortness of breath.

The dynamics of the outbreak was largely shaped by the presence of the so-called super spread event (SSE), in which a single patient was shown to spread the virus to a large number of contacts.[1] It is the SSEs that triggered the large scale of SARS pandemic in China. The first SSE patient is a businessman specialized in fishery wholesale. He was treated in three hospitals from January 30, 2003 to February 10, 2003 and along the way infected at least 78 other individuals including hospital staffs, patients, and close relatives and friends.[1] The second SSE individual, who caused the major spread of the disease out of Guangdong, was a business lady, native of Shanxi province. She went to Guangdong for business in late February and become sick while traveling. She went back to her home province and infected eight family members as well as five hospital staffs. The spread continued to Beijing when she decided to seek better treatment in Beijing.[1,2]

The beginning of the global transmission occurred in Metropole Hotel of Hong Kong where a professor of nephrology from a Guangdong hospital stayed during a private visit. Without knowing, the urologist was infected with SARS-CoV a few days before he traveled to Hong Kong. It is later found that he spread the virus to at least 15 other persons in the hotel and in the hospital where he was treated. Among them, five of the hotel contacts continued their international journeys and further transmitted the disease to Vietnam, Singapore, Canada, and other countries. This marks the true beginning of a disastrous worldwide pandemic (http://www.who.int/csr/sars/en/).

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00026-3
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

Table 26.1 Chronological Events of the SARS Outbreaks

Date	Event
November 16, 2002	The first recognized SARS patient, in Foshan, Guangdong province, China
November 16, 2002 to March 10, 2003	11 independent index cases in Foshan, Heyuan, Jiangmen, Zhongshan, Shunde, Guanzhou, Zhaoqing, Shenzhen, Dongguan, China, resulting in more than 50 secondary infections
January 22, 2003	SARS spreading in Guangdong province
March 22, 2003	SARS spreading to Shanxi and Beijing
February 21, 2003	SARS spreading to Hong Kong, marking the beginning of the global pandemic
February 28, 2003	SARS spreading to Vietnam
March 12, 2003	WHO global travel alert of the SARS pandemic
March 14, 2003	SARS spreading to Canada
March 6, 2003	SARS spreading to Singapore
March 17, 2003	WHO established a 9-nation/11-institute international laboratory network
March 24, 2003	Coronavirus was isolated from SARS patient
April 4, 2003	SARS spreading to Philippines
April 12, 2003	Full-length genome of SARS-CoV determined
April 17, 2003	The international laboratory network announced conclusive identification of SARS-CoV as the causative agent
May 23, 2003	Detected SARS coronavirus in market animals
July 5, 2003	WHO removed the last region from the affected list, effectively marking the end of the outbreak
August 7, 2003	WHO reported a total of 8096 cases and 774 death covering the major 2002–2003 outbreaks
September 2003 to April 2004	Outbreaks caused by laboratory incidents in Singapore, Taiwan, and Beijing
December 16, 2003 to January 8, 2004	Four independent SARS cases in Guangdong, causing mild disease with no death

WHO played a key role in the investigation and control of the SARS outbreak from the very beginning. For the first time in history, WHO issued a global travel alert on March 12, 2003, which greatly reduced the rate of long-distance transmission of the disease. On March 17, 2003 WHO established a 9-nation/11-institute SARS network that played a major role in the rapid identification of the causative agent and development of diagnostic tests. Thanks to the international effort co-coordinated by WHO, the SARS

Downloaded From Vet-eBooks.com

outbreaks were effectively under control by July 5, 2003. This was the first powerful demonstration of the kind of devastation a new infectious disease can cause worldwide and the effectiveness of an international organization when it is running at its peak.

Following the major SARS outbreaks of 2003−2004, there were several minor outbreaks with much smaller impacts. Between December 2003 and January 2004, four independent SARS cases were reported in Guangdong, and none of them led to fetal infection or widespread transmission. Subsequent epidemiological tracing revealed that all cases could be linked to civet trading activities.[3] In addition, there were three laboratory outbreaks in September 2003, December 2003, and April 2004 in Singapore, Taiwan, and Beijing, respectively. The most severe outbreak was associated with the incident in Beijing that resulted in a total of nine infection cases with one death. None of the other two laboratory infections resulted in further spread of the virus.[4]

2. SARS Coronavirus

Rapid identification of causative agent in an outbreak caused by unknown pathogen is the key for an effective response. However, in the case of SARS outbreak, this was not the case. Due to the association of nonspecific clinical symptoms associated with SARS patients, several pathogens were initially "identified" as the potential causes of SARS, which included *Chlamydia*, influenza virus, and paramyxovirus.[5] The confusion continued until March, 2003 when three laboratories independently confirmed that a previously unknown coronavirus was the most likely etiological agent of SARS.[6−8]

Coronaviruses are enveloped viruses with the largest single-stranded, positive-sense RNA genomes currently known, ranging in size from 27 to nearly 32 kb in length. Coronaviruses can infect and cause disease in a broad array of avian and mammal species, including humans. The name "coronavirus" is derived from the Greek word, meaning crown, as the virus envelope appears under electron microscopy to be crowned by a characteristic ring of small bulbous structures. Within the virion, the ssRNA genome is encased in a helical nucleocapsid composed of many copies of the nucleocapsid (N) protein. The lipid bilayer envelope contains three proteins: envelope (E) and membrane (M) protein, which coordinate virion assembly and release, and the large spike (S) protein, which confers the virus's characteristic corona shape as well as serves as the principal mediator of host cell attachment and entry via virus- and host-specific cell receptors. The size of the SARS-CoV viral particle is approximately 80−90 nm and its genomic size is around 29.7 kb.[9,10] The SARS-CoV genome contains 14 open reading frames (ORFs) flanked by 5′- and 3′-untranslated regions of 265 and 342 nucleotides, respectively. While all CoVs carry strain-specific accessory proteins encoded by their downstream ORFs, the order of essential genes—the replicase/transcriptase gene, S gene, E gene, M gene, and N gene—is highly conserved.[11] Similar to other known coronaviruses, the SARS-CoV genome expression starts with two long open reading frames (ORFs), ORF1a and ORF1b, which account for two-thirds of the genomic capacity, followed by ORFs encoding S, E, M, and N proteins (Fig. 26.1). In addition to these conserved core genes in coronaviruses, the SARS-CoV

Downloaded From Vet-eBooks.com

genome contains several accessory genes that are specific to SARS-CoV and their encoded products have no homologue to known proteins. Phylogenetic analysis based on the most conserved gene *ORF1b* indicated that SARS-CoV is distantly related to the group 2 coronaviruses (now the genus *Betacoronavirus*) in the family

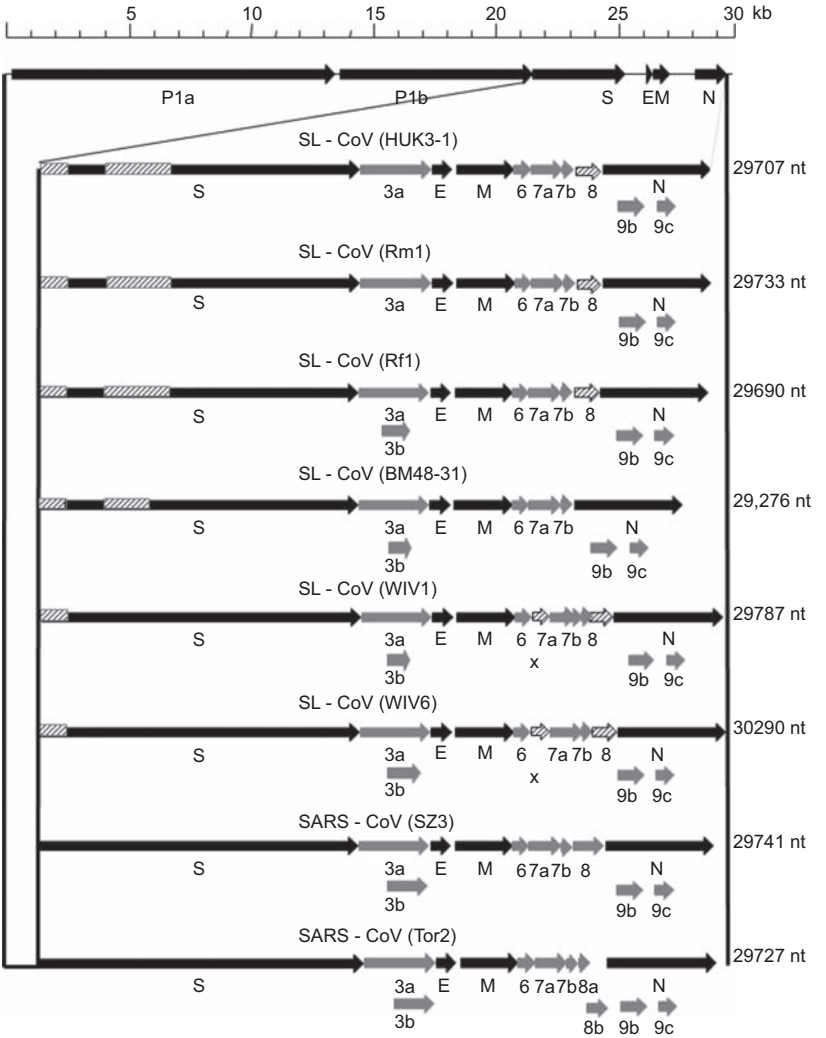


Figure 26.1 Genomic structure of SARS-CoV and bat SL-CoV. The highly conserved genes present in all coronaviruses are shown in *dark-colored arrows* and the betacoronavirus group b-specific ORFs in *light-colored arrows*. The most variable regions are marked with shaded boxes. Rp3, HKU3-1, WIV1, and WIV16 were identified from *R. sinicus* in China; Rm1 and Rf1 from *Rhinolophus macrotis* and *Rhinolophus ferrumequinum*, respectively, in China; BM48-31 from *Rhinolophus blasii*, in Europe; Tor2 from late-phase patient during 2002−2003 SARS outbreak; SZ from civet during 2002−2003 SARS outbreak. * The host of Rp3 was previously identified as *Rhinolophus pearsoni* and later corrected to be *R. sinicus*.[28]

Downloaded From Vet-eBooks.com

Coronaviridae, and represents a distinct cluster, named group 2b (now the genus *Betacoronavirus* group b; Fig. 26.2).[12,13]

3. The Animal Link

Due to the rapid spread of the disease and the delay in the identification of the causative agent, there was no detailed epidemiological tracing done at the beginning of the

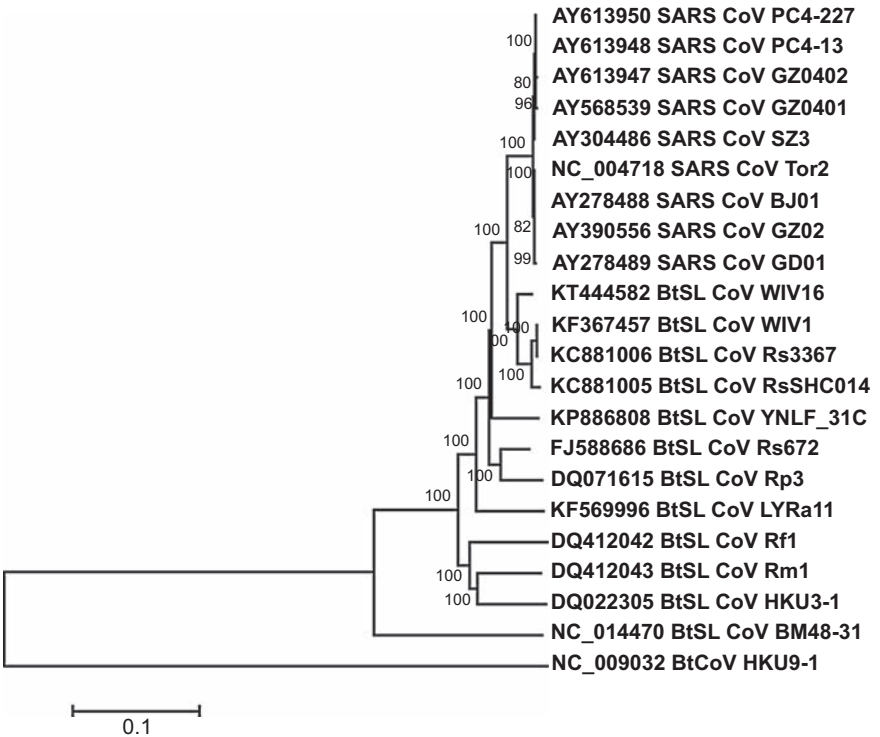


Figure 26.2 Phylogenetic tree of betacoronavirus group b. The phylogenetic tree is generated based on full-length genome sequences of selected SARS-CoVs and bat SL-CoVs using the Neighbor-Joining algorithm in the MEGA4 program[78] with a bootstrap of 1000 replicates. A bat coronavirus BtCoV HKU9 is used as an outgroup.[79] Numbers above branches indicate bootstrap values from 1000 replicates. Scale bar, 0.5 substitutions per site. GD01: SARS-CoV isolate from early-phase patient during 2002−2003 SARS outbreak; Tor2, BJ01: SARS-CoV isolate from late-phase patient during 2002−2003 SARS outbreak; SZ: SARS-CoV isolate from civet during 2002−2003 SARS outbreak; GZ0401/02: SARS-CoV isolate from patient during 2003−2004 SARS outbreak; and PC4-13, PC4-227: SARS-CoV isolate from civet during 2003−2004 SARS outbreak. BtSL-CoV: bat SARS-like CoV. Rp3, HKU3-1, WIV, WIV16, and LYRa11 were identified from *R. sinicus* in China; Rm1 from *Rhinolophus macrotis* in China; Rf1 and YNLC31 C from *Rhinolophus ferrumequinum* in China; and BM48-31 from *Rhinolophus blasii*, in Europe.

Downloaded From Vet-eBooks.com

outbreaks, and it was therefore impossible to trace the origin of the virus. However, through retrospective investigation, it emerged that the majority of the early index cases were limited in several cities of the Guangdong province and most of them have history of contact directly or indirectly with wildlife animals, including handling, killing, and selling wildlife animals as well as preparing and serving wildlife animal meat in restaurants.[14–16]

As these epidemic regions have a unique dietary tradition favoring freshly slaughtered game meat, there is a huge trafficking and trading industry dedicated to live animal trading in specialized market, the "wet market." Immediately after SARS-CoV was identified as the etiological agent of SARS, studies were conducted in those markets for evidence of SARS-CoV in market animals. One of the earliest and most important studies was conducted by a joint team from Hong Kong and Shenzhen in mainland China.[14] In this investigation, out of 25 samples collected from market animals, SARS-CoV-like viruses were isolated from four out of six masked palm civets (*Paguma larvata*) and one raccoon dog (*Nyctereutes procyonoides*). Antibodies against SARS-CoV were detected in masked palm civets, raccoon dog, and Chinese ferret-badgers (*Melogale moschata*). Genome sequencing indicated that the viruses isolated from civets were almost identical to those from human, suggesting a highly possible zoonotic transmission of SARS-CoV from animal(s) to human.[14] These data indicated that at least three different animal species were infected by a coronavirus that is closely related to SARS-CoV. This important study provided the first direct evidence that SARS-CoV existed in animals, pointing to an animal link of the SARS outbreaks.

Although three animals were identified as susceptible to SARS-CoV infection, the larger sale volume of civets in comparison to other animals in the market made them the target animals of subsequent surveillance studies. The role of civets as a major carrier of SARS-CoV in the markets was further confirmed by serological studies involving much large samples.[17,18]

The most detailed epidemiological data proving a direct civet to human transmission of SARS-CoV was obtained during the investigation of the second wave of SARS outbreaks during December 2003 to January 2004. There were two lines of evidences suggesting a direct transmission. First, all four independent cases had the history of direct or indirect contact with civets. Second, sequencing analysis indicated that sequences derived from human samples were more closely related to those in the civets during that period than those from human samples obtained in the major 2002–2003 outbreaks.[3]

In summary, based on the previously mentioned study findings, it was concluded that the civet to human transmission is a major, if not the only, source of SARS-CoV introduction into the human population.[19–21]

4. Natural Reservoirs of SARS-CoV

Natural reservoir refers to the long-term host of the pathogen of an infectious disease. It is often the case that hosts do not get the disease carried by the pathogen, or the

Downloaded From Vet-eBooks.com

infection in the reservoir host is subclinical, asymptomatic, and nonlethal. Once discovered, natural reservoirs elucidate the complete life cycle of infectious diseases, which in turn will help to provide effective prevention and control strategies.

As stated earlier, it is clear that civets played a pivotal role in the 2002−2004 outbreaks of SARS in southern China. Culling of civets seemed to be effective in controlling further outbreaks in the region. However, the role of civets as a potential natural reservoir host was less evident and eventually ruled out by several studies. Serological and molecular studies indicated that only civets in the markets were infected with SARS-CoV whereas the populations of civets in the wild or on farms were free of major infections.[18,22,23] Civets produced overt clinical syndromes when experimentally infected with SARS-CoV.[24] Comparative genome sequence analysis indicated that SARS-CoVs in civets experienced rapid mutation, suggesting that the viruses were still adapting to the host rather than persisting in equilibrium expected for viruses in their natural reservoir species.[17,25]

Continuing search for the potential reservoir host of SARS-CoV resulted in the simultaneous discovery of SARS-like coronaviruses (SL-CoVs) in bats by two independent teams in 2005. Using serological and PCR surveillance, both groups discovered that SL-CoVs were present in different horseshoe bats in the genus *Rhinolophus*.[22,26] Complete genome sequence analysis revealed that bat SL-CoVs have an identical genome organization and a nucleotide sequence identity of 88−92% to SARS-CoV (Fig. 26.1; Table 26.2). Except for the S, ORF3, and ORF8 gene products, all deduced aa sequences of the other gene products have a sequence identity above 93% with those of SARS-CoV. The variable regions between SARS-CoV and bat SL-CoV are mainly located in the coding regions for the nonstructural protein 3 (Nsp 3), S protein, ORF3, and ORF8, the products of these genes have aa sequence identity of 87−95%, 76−78%, 82−90%, and 34−80%, respectively. Among the different bat SL-CoVs, the coding regions for these proteins also represent the most variable regions.[27−29]

The phylogenetic analysis indicated that bat SL-CoVs were grouped in the same cluster of SARS-CoV and were only distantly related to other previously known coronaviruses (Fig. 26.2). To date, these bat SL-CoVs represent naturally occurring CoVs that are most closely related to the SARS-CoVs isolated from humans and civets.

Analysis of nonsynonymous and synonymous substitution rates in bat SL-CoVs suggests that these viruses are not experiencing a positive selection pressure that would be expected if horseshoe bats are new host to these viruses. Instead, these data would argue that these viruses have been associated with the bat hosts for a long time.[27,29,30] These observations would support the notion that bats in the genus *Rhinolophus* are the likely natural reservoir hosts of bat SL-CoVs. It can be further postulated that similar bat species may serve as natural reservoirs of viruses with closer evolutionary relationship to the viruses that were responsible for the 2002−2004 SARS outbreaks.

In this context, we and other groups continued the search for the direct progenitor of SARS-CoV and made great progress in the last 10 years following the initial discovery of SL-CoVs in horseshoe bats. First, highly diverse SL-CoVs have been found not only in Chinese but also in European and African bats, indicating a much wider geographic distribution and long evolutionary history of SL-CoVs in different bat populations

Downloaded From Vet-eBooks.com

Table 26.2 Comparison of Gene Products Between SARS-CoV and Bat SL-CoV

	Gene Product Size (aa)							Amino Acid Sequence Identity With Tor2/sz3 (%)[a]				
Gene/ORF	**Tor2**	**SZ3**	**Rf1**	**Rp3**	**Rm1**	**HKU3-1**	**Rs1**	**Rf1**	**Rp3**	**Rm1**	**HKU3-1**	**Rs672**
P1a	4382	4382	4377	4380	4388	4376	4189	94	96	93	94	94
P1b	2628	2628	2628	2628	2628	2628	2628	98	99	98	98	99
nsp3[b]	1922	1922	1917	1920	1928	1916	1729	92	95	90	92	87
S	1255	1255	1241	1241	1241	1242	1241	76	78	78	78	79
S1	680	680	666	666	666	667	666	63	63	64	6	64
S2	575	575	575	575	575	575	575	92	96	96	94	96
ORF3a	274	274	274	274	274	274	274	86	83	83	82	90
ORF3b	154	154	113	56	56	39	114	89	NA	NA	NA	97
ORF3c	NP	NP	32	NP	NP	NP	NP	NA	NA	NA	NA	NA
E	76	76	76	76	76	76	76	96	100	98	100	100
M	221	221	221	221	221	221	221	97	97	97	99	99
ORF6	63	63	63	63	63	63	63	93	92	92	94	98
ORF7a	122	122	122	122	122	122	122	91	95	93	94	96

Downloaded From Vet-eBooks.com

ORF7b	44	44	44	44	44	44	44	90	93	93	93	93
ORF8a	39	NP	NP	NP	NP	NP	NP	NA	NA	NA	NA	NA
ORF8b	84	NP	NP	NP	NP	NP	NP	NA	NA	NA	NA	NA
ORF8	NP	122	122	121	121	121	121	80	35	35	34	36
N	422	422	421	421	420	421	422	95	97	97	96	99
ORF9a	98	98	96	97	97	97	98	81	85	90	88	92
ORF9b	70	70	70	70	70	70	70	80	91	91	88	94

NP, not present; *NA*, not applicable.
[a]Tor2 was used for all homology calculations with the exception of ORF8, which is absent in Tor2, the SZ3 was used instead.
[b]The region of nsp3 is highly variable and was calculated alone.

Downloaded From Vet-eBooks.com

(Table 26.2).[31–37] Second, great genetic diversity of SARS-CoVs were discovered in one particular population of *R. sinicus* in China by a longitudinal surveillance.[31,38] Third and most importantly, two SL-CoV strains were isolated in Vero cells. These two isolates are closely related to the progenitor of the SARS-CoV not only in genomic sequences but also in receptor usage[31,34](Figs. 26.1–26.3; Table 26.1).

5. Molecular Evolution of SARS-CoV in Humans and Animals

Analysis of the large number of SARS-CoV and SL-CoV sequence datasets accumulated since 2004 has clearly demonstrated the importance of virus evolution in cross-species transmission and in pathogenesis. The following is a summary of the major evolutionary findings in host switching, recombination, and virus–receptor interactions.

5.1 Rapid Adaptation of SARS-CoVs in Humans

On the basis of the epidemiological data, the Chinese SARS molecular epidemiology consortium divided the course of the 2002–2004 outbreaks into three stages, the early, middle, and late phases, respectively.[1] The early phase is defined as the period from the first emergence of SARS to the first documented SSE. The middle phase refers to the ensuing events up to the first cluster of SARS cases in a hotel (Hotel M) in Hong Kong, while cases following this cluster fall into the late phase.

Analysis of all the viral sequences available from human patients and animals revealed two major hallmarks of rapid virus evolution during the initial stages of the 2002–2003 outbreaks: (1) All isolates from early patients and market animals contained a 29-nucleotide (nt) sequence in ORF8 that is absent in most of the publicly available human SARS-CoV sequences derived from later phases of the outbreaks; (2) characteristic motif of single-nucleotide variations (SNVs) were identified in SARS-CoVs of different phases and all these SNVs were located in the S gene that codes for the spike protein responsible for attachment to the host cellular receptor.[25] All SARS-CoV isolates from epidemic countries and regions outside mainland China could be traced to Guangdong or Hong Kong based on the S-gene SNV motif.[23,39]

During the second sporadic outbreaks of 2003–2004, it was shown that the SARS-CoV sequences from index patients were almost identical to that from civets collected in the same period and all retained the 29-nt sequence in the ORF8 gene. The mild disease symptoms associated with these viruses and the lack of rapid human-to-human transmission provided further evidence that the rapid adaptation of the SARS-CoV in the first major outbreak of 2002–2003 was essential for its establishment and pathogenesis in humans.

With the available genomic variation data and the sampling time, it is now possible to calculate the neutral mutation rate and to estimate the date for the most recent common ancestors (MRCAs) of SARS-CoV. The estimate obtained is around 8.00×10^{-6} nt^{-1} day^{-1}, suggesting that SARS-CoV evolves at a relatively constant

Downloaded From Vet-eBooks.com

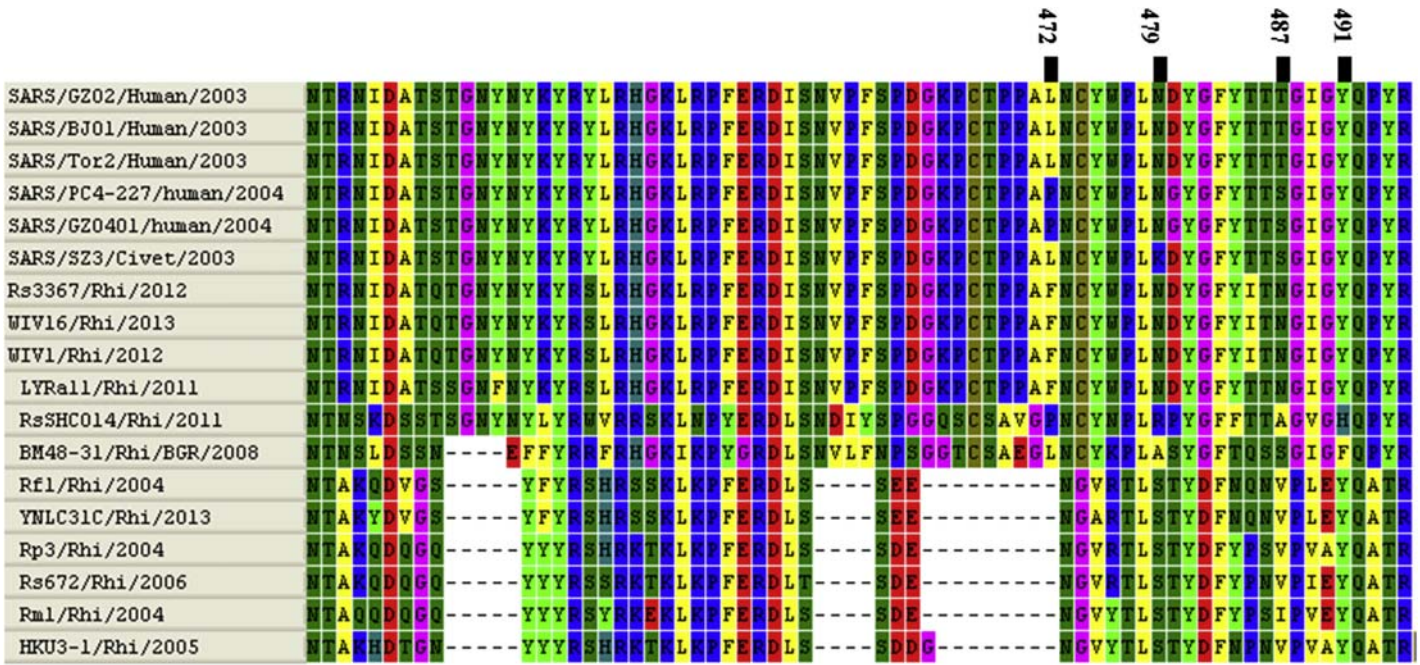


Figure 26.3 Alignment of amino acid sequences covering the receptor-binding motif from viruses of different species origin. GD01: SARS-CoV isolate from early-phase patient during 2002−2003 SARS outbreak; Tor2, BJ01: SARS-CoV isolate from late-phase patient during 2002−2003 SARS outbreak; SZ: SARS-CoV isolate from civet during 2002−2003 SARS outbreak; GZ0402: SARS-CoV isolate from patient during 2003−2004 SARS outbreak; and PC4-227: SARS-CoV isolate from civet during 2003−2004 SARS outbreak. * indicates the two key residues 479 and 487. Rp3, HKU3-1, WIV, WIV16, and LYRa11 were identified from *R. sinicus* in China; Rm1 from *Rhinolophus macrotis* in China; Rf1 and YNLC31C from *Rhinolophus ferrumequinum* in China; and BM48-31 from *Rhinolophus blasii*, in Europe.

Downloaded From Vet-eBooks.com

neutral rate both in humans and palm civet. From these calculations, it was estimated that the MRCAs for palm civet and humans of different transmission lineages lie in mid-November 2002. This estimate was consistent with the first observed SARS case around November 16, 2002 in Foshan, Guangdong.[1,2,25]

5.2 Generation of Viral Genetic Diversity by Recombination

At the present time, at least 33 full-length genome sequences of bat SL-CoVs were determined.[22,26–29,32,34–37,40] Shown in Fig. 26.1 is a comparison of the genome structures for seven selected bat SL-CoVs and one each of civet and human SARS-CoV isolates. All bat SL-CoVs, with the exception of HKU3-8[29] and BM48-31,[32] contain the 29-nt sequence in ORF8, which is present in SARS-CoV from early-phase patients and civets, indicating the common ancestor between civet SARS-CoV and bat SL-CoV. The SL-CoV HKU3-8 contained a 26-nt deletion that is located 14 nt downstream from the commonly observed 29-nt deletion, and the BM48-31 completely lost the ORF8, indicating that the ORF8 coding region is a "hotspot" for deletions.

SL-CoVs from different bat species share 88–97% nt identity among themselves, indicating that the genetic diversity of SL-CoVs in bats is much greater than that observed among civet or human isolates. The most dramatic sequence difference between human SARS-CoV and bat SL-CoV is in the S protein that has 76–97% aa identity for the whole S protein and 64–95% aa identity for the N-terminal region (or the S1 region; Table 26.2). This great genetic diversity observed among bat SL-CoVs and the major difference between the S1 regions of SL-CoV and SARS-CoV S proteins clearly demonstrated that bats are natural reservoirs of human SARS-CoV.

It is well documented that the positive-sense ssRNA genomes of coronaviruses are prone to homologous recombination during coinfection of different coronaviruses and that recombination plays an important role in generating new coronavirus species, in facilitating cross-species transmission and in modulating virus virulence.

Several studies provided evidence for coinfection and recombination came from analysis of SL-CoVs in bats.[29,41–44] It was further revealed that recombination can occur at multiple sites along the SL-CoV genome.[11,28,29,31,34,41] For example, detailed sequence analysis of two genotypes of bat SL-CoV, Rp3 and Rs672 (both were identified from *R. sinicus*), suggested that they may represent a recombinant of two bat SL-CoVs and one of them is more closely related to the human SARS-CoVs.[28,41] During 2015 and 2016, two teams reported a full-length ORF8 that shares higher sequence similarities to the SARS-CoV GZ02 and civet SARS-CoV SZ3 than previously detected SL-CoVs.[37,40] These results suggest that SARS-CoV most likely originated from different bat SL-CoVs via a complicated evolutionary path that involved recombination events.

5.3 Receptor Usage and Evolutionary Selection

The S protein of coronavirus is responsible for attachment to cellular receptor to initiate the first step of virus infection. The angiotensin-converting enzyme 2 (ACE2) was identified as a main functional receptor for SARS-CoV.[45] Further analysis demonstrated that the region covering aa 318–520 of S protein is the key receptor-

Downloaded From Vet-eBooks.com

binding domain (RBD), which is both essential and sufficient to bind the human ACE2 molecule in vitro.[46] Detailed analysis of the crystal structure of the RBD−ACE2 complex revealed that 19 key residues have close contact with the receptor molecule, which are located from aa 424 to 474. This region is termed the receptor-binding motif (RBM).[47]

When the existing epidemiological data was analyzed in combination with the data on infectivity of SARS-CoV isolated in humans at the different phases of the outbreaks and SARS-CoV isolates in civets, a clear correlation could be established between the evolution of the S proteins and virus infectivity. It was observed that the S protein is the fastest evolving protein of SARS-CoV during interspecies transmission from animal to human and in the following phases of human to human transmission. The majority of the mutations are located in the S1 domain (31 of a total of 48 SNVs), particularly in the RBD.[1,46] The interaction analysis between the S proteins of different isolates and the ACE2 molecules demonstrated that two aa residues in the S protein, aa 479 and aa 487, played a key role in virus infectivity.[48,49] For aa residue 479, all 2002−2003 human isolates contain asparagine (N). The palm civet isolates seem to have variable aa residues at this position, all 2002−2003 and some 2003−2004 civet isolates have lysine (K) while other 2003−2004 isolates have either asparagine (N) or arginine (R). For aa residue 487, all isolates including those from early- and middle-phase patients, civets of 2002−2003 and 2003−2004, have a codon for serine (S), whereas all isolates from 2002−2003 late-phase human patients have a codon for threonine (T) (Fig. 26.3). When examined using an HIV-based pseudovirus infection assay, S proteins with all combinations of residues 487/479 could efficiently use the civet ACE2 as an entry receptor, but showed different infectivity in human ACE2-mediated infection.[48,49] The combination of N479/T487 had the highest infectivity, N479/S487 medium infectivity, and K479/S487 the lowest, which almost abolished the infection. These results demonstrated elegantly at the molecular interface that the rapid evolution of the S protein, especially at the aa positions important for host receptor engagement, was essential for the adaptation to and establishment of an effective and productive human infection.

When the genome sequences of SL-CoVs were analyzed, it became evident that the N-terminal regions of their S proteins are the most divergent among themselves, as well as with the SARS-CoV. As shown in Fig. 26.3, bat SL-CoVs can be grouped into three groups based on the RBM sequences. The strains discovered early are close to each other and have a major sequence difference involving deletions of 17−18 aa right in the middle of RBM. We have since demonstrated experimentally that SL-CoV S proteins are unable to use ACE2 molecule, regardless of its origin, as a functional receptor. The second group, identified from European bats, has deletions of 4 aa.[32] The third group, discovered recently, has no deletion and contains an identical size as the SARS-CoV in the S protein (Fig. 26.3).[31,34,35] As predicted from their S sequences, three isolates from the third group, SL-CoV—WIV1, WIV16, and SHC014, have been shown to be able to use ACE2 for cellular entry, even though these S proteins still have slight difference at the key aa involved in direct interaction with ACE2.[31,34,50] Most importantly, the SHC014 can replicate well in transgenic mice containing human ACE2, and it caused tissue damage in tested animals.[50]

Downloaded From Vet-eBooks.com

6. Coronavirus Surveillance in Wildlife Animals

Zoonosis contributes to the majority of emerging disease in the last 30 years, many of them originated from wildlife animals.[51–55] The story of SARS is just one of such examples that spectacularly demonstrated the seamless evolution of a bat virus into a human pathogen responsible for one of the most severe global pandemic outbreaks in modern history of mankind. In general, pathogens carried by wildlife reservoir animals usually do not cause clinical symptoms and they lie dormant until they spill over into and cause diseases in domestic animals or humans. Classical outbreak response measures, such as those deployed during the SARS outbreaks, are still useful, but no longer sufficient for early detection and prevention of major infectious disease outbreaks in the 21st century.

With the demonstration of an increasing number of spillover events that led to severe disease outbreaks in human and domestic animals, we believe it is paramount that from now on we include active surveillance of wildlife animals as part of an integrated infectious disease prevention and control strategy. Surveillance of wildlife animals has also been made more feasible and productive, thanks to the advance in modern molecular techniques including PCR with virus group-specific primers, virus discovery using next generation high-throughput sequencing technologies, and high density virus microarrays.[56–63] Since the SARS outbreaks, especially after the discovery of SL-CoVs in bats, there is a significant surge in international effort for surveillance of coronaviruses in wildlife animals. Before the SARS outbreak, there were only 10 coronaviruses with complete genomes sequenced. This number has increased more than sixfold as a result of the active surveillance works conducted around the world.[27–29,31,32,34,40,55,64–73] Although this only marks the beginning of our understanding of coronaviruses in wildlife animals, it is fair to say that we have learnt a lot more about coronaviruses in general than the past 50 years or so; during that period studying of viruses was only possible and called for in response to disease outbreaks. Based on phylogenetic analysis of the large number of bat coronavirus sequences available presently, it is postulated that all known disease-causing coronaviruses previously identified in humans or animals originated from bat strains.[31,34,43,55] This hypothesis was unfortunately proved by the outbreak of another SARS-like disease, Middle East Respiratory Syndrome (MERS), which was caused by a novel coronavirus (previously named HCoV-EMC, now MERS-CoV) and supposed to originate from bats.[74] Even though the MERS-CoV-like viruses found in bats are not the direct progenitor of the MERS-CoV, the highly genetic diversity of these bat viruses is likely the gene sources for the deadly pathogen in humans, just like that for SARS-CoV.[71,72,75–77]

7. Concluding Remarks

The emergence of SARS-CoV has had a huge impact on the global health and economy. It served as a warning to what may come out of a seemingly harmless virus–reservoir equilibrium in bats or any other wildlife species. At the same time, the experience

Downloaded From Vet-eBooks.com

gained from the SARS outbreaks and the following in-depth studies on SARS-like coronaviruses has provided and will continue to provide invaluable knowledge and guideline to our future fight against new and emerging infectious diseases. One of the major lessons is that we need to pay much more attention to the reservoir species in understanding the genetic diversity of different viruses, the intricate interplay at the virus−host interface, and the major factors responsible for the disturbance of virus−host equilibrium, which in turn trigger spillover events leading to disease outbreaks.

References

1. Chinese SARS Molecular Epidemiology Consortium. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. *Science* 2004;**303**(5664):1666−9.
2. Zhao GP. SARS molecular epidemiology: a Chinese fairy tale of controlling an emerging zoonotic disease in the genomics era. *Philos Trans R Soc Lond B Biol Sci* 2007;**362**(1482):1063−81.
3. Wang M, Yan M, Xu H, Liang W, Kan B, Zheng B, et al. SARS-CoV infection in a restaurant from palm civet. *Emerg Infect Dis* 2005;**11**(12):1860−5.
4. Lim W, Ng KC, Tsang DN. Laboratory containment of SARS virus. *Ann Acad Med Singap* 2006;**35**(5):354−60.
5. WHO. Severe acute respiratory syndrome (SARS). *Wkly Epidemiol Rec* 2003;**78**:81−8.
6. Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, et al. A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* 2003;**348**(20):1953−66.
7. Peiris JS, Lai ST, Poon LL, Guan Y, Yam LY, Lim W, et al. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* 2003;**361**(9366):1319−25.
8. Drosten C, Gunther S, Preiser W, van der Werf S, Brodt HR, Becker S, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med* 2003;**348**(20):1967−76.
9. Marra MA, Jones SJM, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YSN, et al. The genome sequence of the sars-associated coronavirus. *Science* 2003;**300**(5624):1399−404.
10. Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* 2003;**300**(5624):1394−9.
11. Graham RL, Baric RS. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *J Virol* 2010;**84**(7):3134−46.
12. Snijder EJ, Bredenbeek PJ, Dobbe JC, Thiel V, Ziebuhr J, Poon LLM, et al. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J Mol Biol* 2003;**331**(5):991−1004.
13. de Groot R, Baker S, Baric R, Enjuanes L, Gorbalenya A, Holmes K, et al. Family Coronaviridae. In: King A, Adams M, Cartens E, Lefkowitz E, editors. *Virus taxonomy; ninth report of the international committee on taxonomy of viruses*. San Diego, CA: Academic Press; 2012. p. 806−28.
14. Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, et al. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* 2003;**302**(5643):276−8.

Downloaded From Vet-eBooks.com

15. Xu HF, Wang M, Zhang ZB, Zou XZ, Gao Y, Liu XN, et al. An epidemiologic investigation on infection with severe acute respiratory syndrome coronavirus in wild animals traders in Guangzhou. *Zhonghua Yu Fang Yi Xue Za Zhi* 2004;**38**(2):81−3.
16. Xu RH, He JF, Evans MR, Peng GW, Field HE, Yu DW, et al. Epidemiologic clues to sars origin in China. *Emerg Infect Dis* 2004;**10**(6):1030−7.
17. Kan B, Wang M, Jing H, Xu H, Jiang X, Yan M, et al. Molecular evolution analysis and geographic investigation of severe acute respiratory syndrome coronavirus-like virus in palm civets at an animal market and on farms. *J Virol* 2005;**79**(18):11892−900.
18. Tu C, Crameri G, Kong X, Chen J, Sun Y, Yu M, et al. Antibodies to SARS coronavirus in civets. *Emerg Infect Dis* 2004;**10**(12):2244−8.
19. Wang LF, Shi Z, Zhang S, Field H, Daszak P, Eaton BT. Review of bats and SARS. *Emerg Infect Dis* 2006;**12**(12):1834−40.
20. Wang LF, Eaton BT. Bats, civets and the emergence of SARS. *Curr Top Microbiol Immunol* 2007;**315**:325−44.
21. Shi Z, Hu Z. A review of studies on animal reservoirs of the SARS coronavirus. *Virus Res* 2008;**133**(1):74−87.
22. Poon LL, Chu DK, Chan KH, Wong OK, Ellis TM, Leung YH, et al. Identification of a novel coronavirus in bats. *J Virol* 2005;**79**(4):2001−9.
23. Lan YC, Liu TT, Yang JY, Lee CM, Chen YJ, Chan YJ, et al. Molecular epidemiology of severe acute respiratory syndrome-associated coronavirus infections in Taiwan. *J Infect Dis* 2005;**191**(9):1478−89.
24. Wu DL, Tu CC, Xin C, Xuan H, Meng QW, Liu YG, et al. Civets are equally susceptible to experimental infection by two different severe acute respiratory syndrome coronavirus isolates. *J Virol* 2005;**79**(4):2620−5.
25. Song HD, Tu CC, Zhang GW, Wang SY, Zheng K, Lei LC, et al. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc Natl Acad Sci USA* 2005;**102**(7):2430−5.
26. Lau SK, Woo PC, Li KS, Huang Y, Tsoi HW, Wong BH, et al. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc Natl Acad Sci USA* 2005; **102**(39):14040−5.
27. Ren W, Li W, Yu M, Hao P, Zhang Y, Zhou P, et al. Full-length genome sequences of two SARS-like coronaviruses in horseshoe bats and genetic variation analysis. *J Gen Virol* 2006;**87**(Pt 11):3355−9.
28. Zhang Y, Zhang H, Dong X, Yuan J, Yang X, Zhou P, et al. Hantavirus outbreak associated with laboratory rats in Yunnan, China. *Infect Genet Evol* 2010;**10**(5):638−44.
29. Lau SK, Li KS, Huang Y, Shek CT, Tse H, Wang M, et al. Ecoepidemiology and complete genome comparison of different strains of severe acute respiratory syndrome-related Rhinolophus bat coronavirus in China reveal bats as a reservoir for acute, self-limiting infection that allows recombination events. *J Virol* 2010;**84**(6):2808−19.
30. Tang X, Li G, Vasilakis N, Zhang Y, Shi Z, Zhong Y, et al. Differential stepwise evolution of SARS coronavirus functional proteins in different host species. *BMC Evol Biol* 2009;**9**:52.
31. Ge XY, Li JL, Yang XL, Chmura AA, Zhu G, Epstein JH, et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* 2013; **503**(7477):535−8.
32. Drexler JF, Gloza-Rausch F, Glende J, Corman VM, Muth D, Goettsche M, et al. Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. *J Virol* 2010;**84**(21):11336−49.

Downloaded From Vet-eBooks.com

33. Tong S, Conrardy C, Ruone S, Kuzmin IV, Guo X, Tao Y, et al. Detection of novel SARS-like and other coronaviruses in bats from Kenya. *Emerg Infect Dis* 2009;**15**(3):482−5.
34. Yang XL, Hu B, Wang B, Wang MN, Zhang Q, Zhang W, et al. Isolation and characterization of a novel bat coronavirus closely related to the direct progenitor of SARS coronavirus. *J Virol* 2015;**90**(6):3253−6.
35. He B, Zhang Y, Xu L, Yang W, Yang F, Feng Y, et al. Identification of diverse alphacoronaviruses and genomic characterization of a novel severe acute respiratory syndrome-like coronavirus from bats in China. *J Virol* 2014;**88**(12):7070−82.
36. Yang L, Wu Z, Ren X, Yang F, He G, Zhang J, et al. Novel SARS-like betacoronaviruses in bats, China, 2011. *Emerg Infect Dis* 2013;**19**(6):989−91.
37. Wu Z, Yang L, Ren X, Zhang J, Yang F, Zhang S, et al. ORF8-Related genetic evidence for Chinese horseshoe bats as the source of human severe acute respiratory syndrome coronavirus. *J Infect Dis* 2016;**213**(4):579−83.
38. Wang M-N, Zhang W, G Y-T, H B, G X-Y, Y X-L, et al. Longitudinal surveillance of SARS-like coronaviruses in bats by quantitative real-time PCR. *Virol Sin* 2016;**31**(1): 78−80.
39. Tang JW, Cheung JL, Chu IM, Ip M, Hui M, Peiris M, et al. Characterizing 56 complete SARS-CoV S-gene sequences from Hong Kong. *J Clin Virol* 2007;**38**(1):19−26.
40. Lau SK, Feng Y, Chen H, Luk HK, Yang WH, Li KS, et al. Severe acute respiratory syndrome (sars) coronavirus ORF8 protein is acquired from sars-related coronavirus from greater horseshoe bats through recombination. *J Virol* 2015;**89**(20):10532−47.
41. Hon CC, Lam TY, Shi ZL, Drummond AJ, Yip CW, Zeng F, et al. Evidence of the recombinant origin of a bat severe acute respiratory syndrome (SARS)-like coronavirus and its implications on the direct ancestor of SARS coronavirus. *J Virol* 2008;**82**(4):1819−26.
42. Cui J, Han N, Streicker D, Li G, Tang X, Shi Z, et al. Evolutionary relationships between bat coronaviruses and their hosts. *Emerg Infect Dis* 2007;**13**(10):1526−32.
43. Vijaykrishna D, Smith GJ, Zhang JX, Peiris JS, Chen H, Guan Y. Evolutionary insights into the ecology of coronaviruses. *J Virol* 2007;**81**(8):4012−20.
44. Tang XC, Zhang JX, Zhang SY, Wang P, Fan XH, Li LF, et al. Prevalence and genetic diversity of coronaviruses in bats from China. *J Virol* 2006;**80**(15):7481−90.
45. Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 2003;**426**(6965): 450−4.
46. Wong SK, Li W, Moore MJ, Choe H, Farzan M. A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. *J Biol Chem* 2004; **279**(5):3197−201.
47. Li F, Li W, Farzan M, Harrison SC. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* 2005;**309**(5742):1864−8.
48. Qu XX, Hao P, Song XJ, Jiang SM, Liu YX, Wang PG, et al. Identification of two critical amino acid residues of the severe acute respiratory syndrome coronavirus spike protein for its variation in zoonotic tropism transition via a double substitution strategy. *J Biol Chem* 2005;**280**(33):29588−95.
49. Li W, Zhang C, Sui J, Kuhn JH, Moore MJ, Luo S, et al. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J* 2005;**24**(8):1634−43.
50. Menachery VD, Yount Jr BL, Debbink K, Agnihothram S, Gralinski LE, Plante JA, et al. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat Med* 2015;**21**(12):1508−13.
51. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. *Nature* 2008;**451**(7181):990−3.

Downloaded From Vet-eBooks.com

52. Chomel BB, Belotto A, Meslin FX. Wildlife, exotic pets, and emerging zoonoses. *Emerg Infect Dis* 2007;**13**(1):6–11.
53. Bengis RG, Leighton FA, Fischer JR, Artois M, Morner T, Tate CM. The role of wildlife in emerging and re-emerging zoonoses. *Rev Sci Tech* 2004;**23**(2):497–511.
54. Woolhouse ME, Gowtage-Sequeria S. Host range and emerging and reemerging pathogens. *Emerg Infect Dis* 2005;**11**(12):1842–7.
55. Liang YZ, Wu LJ, Zhang Q, Zhou P, Wang MN, Yang XL, et al. Cloning, expression, and antiviral activity of interferon beta from the Chinese microbat. *Myotis Davidii Virol Sin* 2015;**30**(6):425–32.
56. Ng TF, Manire C, Borrowman K, Langer T, Ehrhart L, Breitbart M. Discovery of a novel single-stranded DNA virus from a sea turtle fibropapilloma by using viral metagenomics. *J Virol* 2009;**83**(6):2500–9.
57. Yanai-Balser GM, Duncan GA, Eudy JD, Wang D, Li X, Agarkova IV, et al. Microarray analysis of *Paramecium bursaria* chlorella virus 1 transcription. *J Virol* 2010;**84**(1): 532–42.
58. Breitbart M, Hewson I, Felts B, Mahaffy JM, Nulton J, Salamon P, et al. Metagenomic analyses of an uncultured viral community from human feces. *J Bacteriol* 2003;**185**(20): 6220–3.
59. Gaynor AM, Nissen MD, Whiley DM, Mackay IM, Lambert SB, Wu G, et al. Identification of a novel polyomavirus from patients with acute respiratory tract infections. *PLoS Pathog* 2007;**3**(5):e64.
60. Ge X, Li Y, Yang X, Zhang H, Zhou P, Zhang Y, et al. Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. *J Virol* 2012;**86**(8):4620–30.
61. Lipkin WI, Anthony SJ. Virus hunting. *Virology* 2015;**479–480**:194–9.
62. Briese T, Kapoor A, Mishra N, Jain K, Kumar A, Jabado OJ, et al. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *Mbio* 2015;**6**(5):e01491–515.
63. Barzon L, Lavezzo E, Militello V, Toppo S, Palu G. Applications of next-generation sequencing technologies to diagnostic virology. *Inter J Mol Sci* 2011;**12**(11):7861–84.
64. Vijgen L, Keyaerts E, Lemey P, Maes P, Van Reeth K, Nauwynck H, et al. Evolutionary history of the closely related group 2 coronaviruses: porcine hemagglutinating encephalomyelitis virus, bovine coronavirus, and human coronavirus OC43. *J Virol* 2006;**80**(14): 7270–4.
65. Vijgen L, Keyaerts E, Moes E, Thoelen I, Wollants E, Lemey P, et al. Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. *J Virol* 2005;**79**(3):1595–604.
66. Woo PC, Lau SK, Chu CM, Chan KH, Tsoi HW, Huang Y, et al. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J Virol* 2005;**79**(2):884–95.
67. Woo PC, Lau SK, Huang Y, Yuen KY. Coronavirus diversity, phylogeny and interspecies jumping. *Exp Biol Med (Maywood)* 2009;**234**(10):1117–27.
68. Woo PC, Lau SK, Lam CS, Lai KK, Huang Y, Lee P, et al. Comparative analysis of complete genome sequences of three avian coronaviruses reveals a novel group 3c coronavirus. *J Virol* 2009;**83**(2):908–17.
69. Woo PC, Lau SK, Yip CC, Huang Y, Tsoi HW, Chan KH, et al. Comparative analysis of 22 coronavirus HKU1 genomes reveals a novel genotype and evidence of natural recombination in coronavirus HKU1. *J Virol* 2006;**80**(14):7136–45.

Downloaded From Vet-eBooks.com

70. Alekseev KP, Vlasova AN, Jung K, Hasoksuz M, Zhang X, Halpin R, et al. Bovine-like coronaviruses isolated from four species of captive wild ruminants are homologous to bovine coronaviruses, based on complete genomic sequences. *J Virol* 2008;**82**(24): 12422−31.
71. Corman VM, Baldwin HJ, Tateno AF, Zerbinati RM, Annan A, Owusu M, et al. Evidence for an ancestral association of human coronavirus 229E with bats. *J Virol* 2015;**89**(23): 11858−70.
72. Hu B, Ge X, Wang LF, Shi Z. Bat origin of human coronaviruses. *Virol J* 2015;**12**(1):221.
73. Woo PC, Lau SK, Lam CS, Lau CC, Tsang AK, Lau JH, et al. Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. *J Virol* 2012;**86**(7): 3995−4008.
74. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *New Engl J Med* 2012; **367**(19):1814−20.
75. Yang Y, Liu C, Du L, Jiang S, Shi Z, Baric RS, et al. Two mutations were critical for bat-to-human transmission of Middle East respiratory syndrome coronavirus. *J Virol* 2015;**89**(17): 9119−23.
76. Wang Q, Qi J, Yuan Y, Xuan Y, Han P, Wan Y, et al. Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26. *Cell Host Microbe* 2014;**16**(3): 328−37.
77. Lau SK, Li KS, Tsang AK, Lam CS, Ahmed S, Chen H, et al. Genetic characterization of Betacoronavirus lineage C viruses in bats reveals marked sequence divergence in the spike protein of pipistrellus bat coronavirus HKU5 in Japanese pipistrelle: implications for the origin of the novel Middle East respiratory syndrome coronavirus. *J Virol* 2013;**87**(15): 8638−50.
78. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 2007;**24**(8):1596−9.
79. Lau SK, Poon RW, Wong BH, Wang M, Huang Y, Xu H, et al. Coexistence of different genotypes in the same bat and serological characterization of Rousettus bat coronavirus HKU9 belonging to a novel Betacoronavirus subgroup. *J Virol* 2010;**84**(21):11385−94.

Downloaded From Vet-eBooks.com

This page intentionally left blank

Downloaded From Vet-eBooks.com

Ecology and Evolution of Avian Influenza Viruses

27

A.C. Hurt[1,2], R.A.M. Fouchier[3], D. Vijaykrishna[4]
[1]WHO Collaborating Centre for Reference and Research on Influenza, VIDRL, at the Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia; [2]University of Melbourne, Parkville, VIC, Australia; [3]Erasmus Medical Center, Rotterdam, the Netherlands; [4]Duke-NUS Graduate Medical School, Singapore, Singapore

1. Introduction to Influenza A Virus

1.1 Taxonomy and Host Range

Influenza A viruses belong to the family *Orthomyxoviridae*, a family of viruses with a negative sense, single-stranded, segmented RNA genome.[1,2] Other members of the family include the human pathogens—influenza B virus and influenza C virus; a newly identified genus in cows, which has been tentatively classified as influenza D virus[3]; the tick-borne viruses in the *Thogotovirus* genus; the infectious salmon anemia virus within the *Isavirus* genus; and viruses detected in ticks, birds, and humans of the *Quaranjavirus* genus.[4] Influenza A viruses have been isolated from many host species including humans, pigs, horses, mink, cats, dogs, marine mammals, and a wide range of domestic birds, but wild birds in the orders *Anseriformes* (ducks, geese, and swans) and *Charadriiformes* (gulls, terns, and waders) are thought to form the virus reservoirs in nature.[5] Virus spillover from the reservoir hosts is sporadic, although the continuous circulation and persistence of introduced viruses have been observed in poultry, dogs, horses, pigs, and humans (Fig. 27.1).

1.2 Influenza A Virus Structure and Genome Organization

The influenza A virus genome consists of eight gene segments (Fig. 27.2).[1,2] Segments 1−3 encode the polymerase proteins: basic polymerase 2 (PB2), basic polymerase 1 (PB1), and acidic polymerase (PA), respectively. Segment 2 also encodes two further proteins that are expressed independently via a second open reading frame, PB1-F2, which has been implicated in the induction of cell death,[6,7] and PB1-N40.[8] Segment 3 also contains a second open reading frame that expresses PA-X, a protein which modulates host response to influenza infection.[9] Segments 4 and 6 encode the viral surface glycoproteins hemagglutinin (HA) and neuraminidase (NA). HA is responsible for binding to sialic acids, the viral receptors on host cells, and for fusion of the viral and host cell membranes upon endocytosis. NA is a sialidase, responsible for cleaving sialic acids from virus-producing host cells and virus particles, thus facilitating virus

Genetics and Evolution of Infectious Diseases. http://dx.doi.org/10.1016/B978-0-12-799942-5.00027-5
Copyright © 2017 Elsevier Inc. All rights reserved.

Downloaded From Vet-eBooks.com

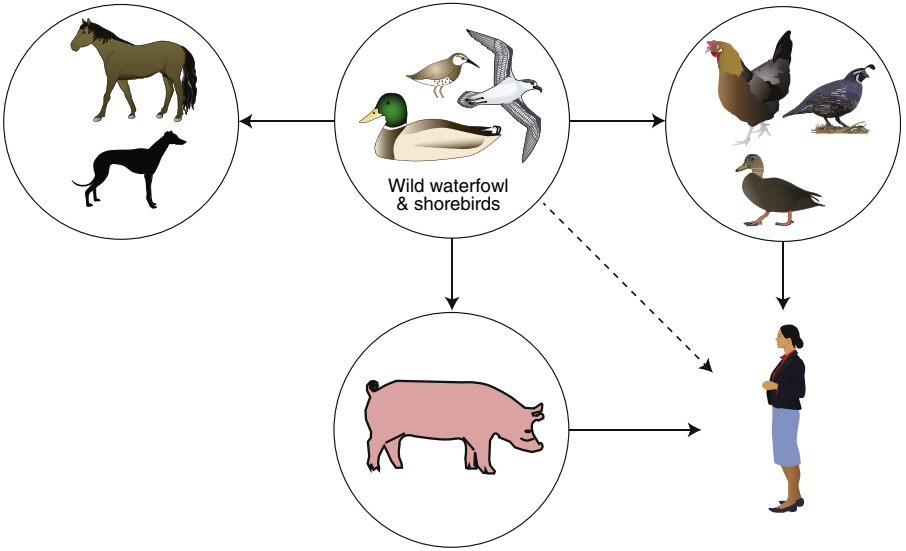


Figure 27.1 Wild waterfowl and shorebirds form the reservoir for influenza A viruses in nature. Avian viruses are occasionally transmitted to other hosts, in which they may cause serious disease. Only major routes of transmission are indicated with *arrows*, and in this chapter, only influenza viruses of birds—wild and domestic—are discussed.

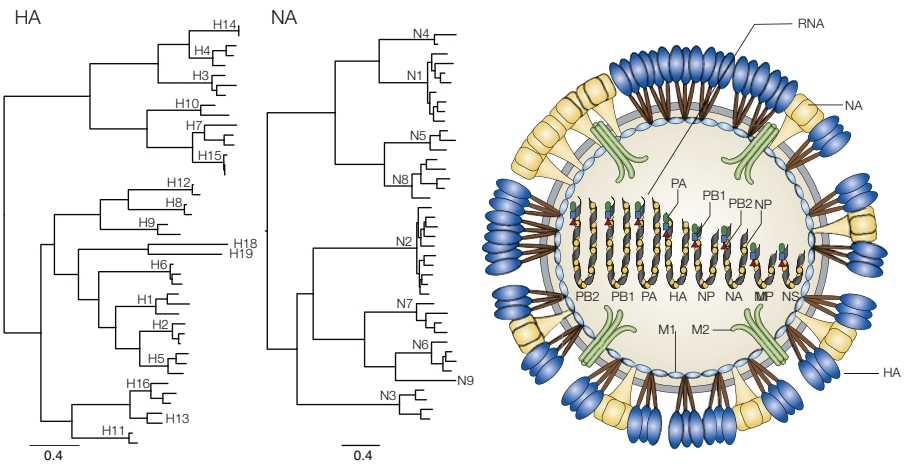


Figure 27.2 Influenza A virus and genetic variation of the surface glycoproteins HA and NA. Phylogenetic trees demonstrate the genetic variation of the surface glycoprotein genes HA and NA in the wild bird reservoir, and in the case of HA, include the newly detected H17 and H18 viruses from bats. A schematic presentation of an influenza A virus particle with its eight RNA gene segments and virus-associated proteins (named).
Adapted from Karlsson Hedestam GB, Fouchier RA, Phogat S, Burton DR, Sodroski J, Wyatt RT. The challenges of eliciting neutralizing antibodies to HIV-1 and to influenza virus. *Nat Rev Microbiol* 2008;**6**(2):143−155.

Downloaded From Vet-eBooks.com

release. Segment 5 codes for the nucleocapsid protein (NP) that binds to viral RNA and together with the polymerase proteins forms the ribonucleoprotein complexes. Segment 7 codes for the classical viral matrix structural protein M1, and the ion channel protein M2 that is incorporated into the viral membrane. Segment 8 encodes the nonstructural protein NS1 and the nucleic export protein (NEP), previously known as NS2. NS1 is an antagonist of host innate immune responses and interferes with host gene expression while NEP is involved in the nuclear export of RNPs into the cytoplasm before virion assembly.[1,2]

1.3 Influenza A Virus Classification

Influenza A viruses are highly variable. Most pronounced is the genetic and antigenic variation of the surface glycoproteins HA and NA.[5] To date, 16 major antigenic variants of HA and 9 antigenic variants of NA have been detected in birds. Of the possible 144 HA/NA combinations from birds, 127 subtypes have been reported from nature (deduced from sequences submitted to GenBank as of February 25, 2016). In addition to the virus subtypes found in birds, two new influenza A viruses have been detected in bats that have nominally been classified as H17N10 and H18N11.[10,11] However, the neuraminidases of these viruses are distantly related to the N1−N9 NAs of avian influenza and lack the conserved amino acid residues associated with sialic acid binding or cleavage.[12] The classification system is biologically relevant, as host antibodies that recognize one HA or NA subtype will not cross-react, or poorly react, with other HA or NA subtypes. The antigenic variation of the HA and NA proteins is in agreement with the major genetic variation of the respective genes of avian influenza A viruses (Fig. 27.2). For instance, the maximum amino acid sequence identity between the HA of any two different avian influenza HA subtypes (H1−H16) is 79%, and within a subtype can be as low as 86%.[13] The genetic variation of the HA and NA genes in the avian reservoir is of the same order of magnitude as the genetic variation of the surface glycoproteins of primate lentiviruses, a notoriously variable group of viruses.[14]

Besides classification based on the antigenic properties of HA and NA, avian influenza A viruses can be classified based on their pathogenic phenotype in chickens, that is, low pathogenic avian influenza (LPAI) viruses or highly pathogenic avian influenza (HPAI) viruses. The vast majority of avian influenza viruses are LPAI and cause either a mild disease or no symptoms depending on the virus and the species infected. However, two influenza subtypes, H5 and H7, also have the potential to occur as HPAI viruses. HPAI viruses are classified as those that cause more than 75% mortality in chickens following intravenous infection, and are thought to arise in poultry upon introduction of LPAI H5 and H7 viruses from the wild bird reservoir.[15,16] The HA protein of influenza A virus is initially synthesized as a single polypeptide precursor (HA_0), which is cleaved into HA_1 and HA_2 subunits by trypsin-like proteases in the host cell. The switch from the LPAI to the HPAI virus phenotype for H5 and H7 viruses is achieved by the introduction of basic amino acid residues into the HA_0 cleavage site, facilitating cleavage by ubiquitously expressed proteases.[16,17] It is thought that this property facilitates systemic virus replication and a mortality of up to 100% in poultry.

Downloaded From Vet-eBooks.com

2. Influenza Viruses in Birds

2.1 Low Pathogenic Avian Influenza Virus Subtypes in Wild Birds

Avian influenza viruses are most frequently detected from ducks (*Anseriformes*), particularly from dabbling ducks, such as mallards (*Anas platyrhynchos*).[18–28] Typically, infections involve influenza virus subtypes H1–H12 and a wide variety of HA/NA subtype combinations. In addition to *Anseriformes*, extensive surveillance studies have been directed toward wader species within the *Charadriidae* and *Scolopacidae* families, primarily in North America, where LPAI viruses of subtypes H1–H12 have been isolated from waders in eastern USA.[23,29] LPAI viruses have also been detected in waders throughout the world but at lower prevalence than observed in *Anseriformes*.[24,30,31] Gulls and terns appear to represent the H13 and H16 reservoir hosts in nature, as these subtypes are frequently found in these species, but only very rarely in other bird species.[13,24,29,32] In many large surveillance studies, LPAI H14 and H15 viruses are often the only subtypes not detected.[33,34] Although these subtypes are found on rare occasions,[35,36] it suggests that the host reservoir of these viruses may be infrequently sampled in typical studies.

LPAI viruses have been detected in numerous other bird species as well,[25] but it is unclear whether the viruses are truly enzootic in these species or whether these birds are "transient" host species. Such transient host species may share the same habitat for part of the year with species in which LPAI virus prevalence is relatively high, such as ducks. For instance, LPAI viruses are occasionally detected in geese, swans, rails, petrels, cormorants, passerines, and other bird species, although their prevalence is much lower than in dabbling ducks. Due to limitations of wild bird surveillance studies, it cannot be excluded that LPAI virus reservoir species may exist beyond the orders *Anseriformes* and *Charadriiformes*.

2.2 Low Pathogenic Avian Influenza Virus Transmission and Epidemiology in Wild Birds

The prevalence of LPAI viruses in wild birds varies with geographical location, time of year, and bird species.[25] Although spatiotemporal patterns of virus prevalence have been described, significant variations in these patterns may exist from year to year and between different surveillance studies. The seasonal virus prevalence in mallards in North America and Europe may vary from very low ($<1\%$) in spring and summer to very high ($\sim 30\%$) during fall migration and winter.[5,24,28,37] The peak virus prevalence during fall migration is believed to be related to the large numbers of young immunologically naive birds of the breeding season that aggregate before and during southbound migration.[5,38] Indeed, age-related differences in LPAI virus prevalence were detected in mallards and Eurasian wigeons (*Anas penelope*) in Europe, with virus prevalence of 6.8% in juvenile ducks and 2.8% in adults.[24] The peak virus prevalence likely gradually declines when migration proceeds, forming a gradient in LPAI virus prevalence from the breeding grounds of the birds in the north to the wintering areas in

Downloaded From Vet-eBooks.com

the south.[24] It is likely that with increasing age, ducks mount an immune response that limits subsequent infections with LPAI viruses. Upon experimental inoculation of mallards, it has been shown that birds could become reinfected with a heterologous LPAI virus subtype, but with markedly reduced duration of virus shedding. This reduction was more pronounced upon reinfection with an LPAI virus of homologous HA subtype.[39–41] However, under field conditions, consecutive or simultaneous infections with different LPAI virus subtypes are common in dabbling ducks, suggesting that heterosubtypic immunity induced by LPAI viruses is only partial.[42] In wild-caught mallards, it was further estimated that virus shedding occurred for about 10 days,[42] in agreement with the duration of virus shedding under experimental conditions of 7–17 days.[39–41]

Similar to that seen in *Anseriformes*, the highest virus prevalence reported in gulls is late summer, shortly after the introduction of naive juveniles into the population.[43] Most gull species breed in dense colonies, with adults and juveniles crowded in a small space, creating an ideal opportunity for virus spread, explaining why epizootics start within the breeding colony. In contrast, epizootics in ducks are likely initiated when they congregate in large numbers during molting, migration, or wintering.

Interestingly, the patterns of LPAI prevalence were reversed in waders migrating along the Atlantic-America flyway compared to that of ducks. Peak LPAI virus prevalence of about 14% was observed during the spring migration of waders, most notably among ruddy turnstones (*Arenaria interpres*) in Delaware Bay, rather than during the fall migration.[23] This observation led Krauss et al. to hypothesize that waders and ducks could both be important reservoir hosts to maintain the annual cycle of LPAI virus epidemics, in which ducks carry viruses southbound in the fall and waders bring the viruses northbound in spring. While this may be a valid hypothesis based on data from North America, data supporting such an annual cycle elsewhere is not available. Prevalence of LPAI among waders in Europe, Australia, and Alaska is shown to be low, and hot spots for LPAI virus detection equivalent to Delaware Bay have not been observed elsewhere in the world.[24,31,44–46] Undoubtedly, LPAI virus prevalence is also likely to be driven by unique environmental factors in a particular region, including rainfall patterns that influence bird densities,[47] breeding opportunities, and the number of immunologically naive juveniles entering a population, which may impact influenza infection dynamics.[48]

While surveillance studies suggest that LPAI viruses are endemic in dabbling ducks, infection in several other *Anseriformes* species appears more transient. For example, in white-fronted geese in northern Europe, LPAI viruses were only detected following their arrival at the wintering grounds in the Netherlands, suggesting that the birds only became infected after contact with reservoir species, such as mallards.[49] It is possible that many bird species identified in surveillance studies are "transient hosts" while a more limited number of species act as true "reservoir hosts," in which LPAI viruses are considered endemic.

Many LPAI virus host species regularly migrate over long distances. During migration, birds have the potential to distribute viruses between countries or continents. Within the vast continents and along the major flyways, migration connects bird populations in time and space, either at common breeding areas, common foraging areas

Downloaded From Vet-eBooks.com

during migration, or at shared nonbreeding areas. As a result, bird populations may transmit their pathogens to new migratory and nonmigratory populations and to new areas. Virus transmission and geographical spread are thus dependent on the ecology of the migrating hosts. Migrating birds rarely fly the full distance between breeding and nonbreeding areas without stopping and "refueling" along the way. Rather, birds make frequent stopovers during migration and spend more time foraging and preparing for migration than actively performing flights. Many species aggregate at favorable stopover or wintering sites, resulting in high local densities. Such sites may be important for the transmission of viruses between different species of wild birds.[25]

The maintenance and circulation of LPAI viruses within the wild bird host populations rely on effective transmission of the virus between susceptible hosts or host populations. LPAI viruses usually infect cells lining the intestinal tract and are believed to be transmitted primarily via the fecal−oral route.[5] LPAI viruses can stay infectious for prolonged periods of time in surface waters and the environment, potentially allowing viruses to infect various bird populations that might occupy a particular area at a different time.[50−53] Dabbling ducks feed mainly on water surfaces, allowing effective fecal−oral transmission. Diving ducks forage at deeper depths and more often in marine habitats. Dabbling ducks display a propensity for migration and switch breeding grounds between years, in part due to mate choice. These behavioral differences between ecological guilds of ducks could provide an explanation for differences in LPAI virus prevalence in different species.

Given the relatively short duration of LPAI virus shedding by individual infected birds, the spatial and temporal dynamics of LPAI virus circulation may be explained by the continuous circulation within and between bird flocks, or by viral persistence in abiotic reservoirs, such as lakes. There is insufficient data presently available to determine the relative role of either possibility, although the continuous prevalence of LPAI in species such as dabbling ducks may be sufficient for the year-round perpetuation of viruses in these species without the need for environmental persistence.[27]

2.3 Low Pathogenic Avian Influenza and Highly Pathogenic Avian Influenza Viruses in Domestic Birds

Influenza viruses may infect virtually all species of domestic birds, depending mostly on their direct contact with wild birds and wild bird excretions or indirect contact via human activities. In general, influenza viruses originating from wild birds do not cause serious disease in domestic birds, but may result in decreased egg production, mild respiratory illness, and other mild clinical symptoms. Most LPAI outbreaks have a limited duration and limited geographical scale, although large-scale and continuous outbreaks have been reported for the H9N2 subtype in the eastern hemisphere[15,16] and more recently, in 2015, for the H7N9 subtype in China.[54] There is potential for H5 and H7 LPAI viruses to acquire a multibasic cleavage site upon introduction to domestic poultry, thereby dramatically increasing the pathogenicity of the virus.[17] These HPAI viruses can have a devastating impact on chickens and turkeys, with mortality rates of up to 100%.[15,16] In the 2014−2015 period, HPAI outbreaks have occurred due

Downloaded From Vet-eBooks.com

to H5N1 in Asia, the Middle East, Europe, and Africa; H5N2 in North America, Asia, and Europe; H5N6 in Asia; H5N8 in North America, Europe, and Asia; H7N2 in Australia; H7N3 in Mexico; and H7N7 in Germany and the United Kingdom.[55]

While most HPAI outbreaks have been controlled relatively quickly, the Asian HPAI H5N1 virus lineage has been circulating in poultry in East and Southeast Asia continuously since 1997, resulting in the devastation of the poultry industry in those regions over many years.

The number of poultry involved in the HPAI H5N1 outbreaks is unknown, but is likely to be hundreds of millions.[56] In addition, the virus has continued to cause disease and fatalities in humans, due to zoonotic transmissions (transmissions from birds to humans), in numerous countries. In addition, an LPAI virus of the subtype H7N9 detected in poultry markets in China since 2013 has also resulted in human cases and fatalities.[54] Because the H7N9 virus has low pathogenicity in poultry, as opposed to HPAI H5N1, infection is less obvious, thereby potentially increasing the likelihood that humans may become infected when handling or working with poultry. Fortunately, to date, neither the H5N1 nor the H7N9 viruses have acquired the ability to cause sustained human-to-human transmission.

2.4 Highly Pathogenic Avian Influenza H5N1 and H5NX Virus in Wild Birds

Compared to all other HPAI virus outbreaks, the current epizootic of HPAI H5N1 virus is highly unusual in many regards. The HPAI H5N1 virus is considered endemic in many countries throughout Asia, the Middle East, and Africa, which results in regular poultry outbreaks and occasional zoonotic transmission to humans and other mammals, continuously changing genotypes and spill-back of the virus into wild birds, leading to outbreaks and circulation of the virus in those birds.

The ancestral HPAI H5N1 virus is believed to have originated from a virus circulating in domestic geese in Guandong province, China, in 1996 and introduced in Hong Kong poultry markets in 1997.[57] Previously deemed unlikely, the direct transmission of a purely avian virus into humans during the 1997 Hong Kong outbreak signified a paradigm shift.[58] After the local containment of the HPAI H5N1 virus outbreak, the virus reappeared in 2002 to cause an outbreak in waterfowl and various other bird species in two parks in Hong Kong.[59–61] In 2003, the HPAI H5N1 virus was again transmitted to humans, leading to at least one fatal case. There is little information on the circulation of HPAI H5N1 virus during 1997–2002, although it is believed to have continuously circulated in China during that period.[62] HPAI H5N1 virus resurfaced again in 2003–2004 to spread across a large part of Southeast Asia, including Cambodia, China, Hong Kong, Indonesia, Japan, Laos, Malaysia, South Korea, Thailand, and Vietnam.

Before 2005, HPAI H5N1 viruses had only been isolated sporadically from wild birds, but in April–June 2005, the first reported outbreak in wild migratory birds occurred, in Lake Qinghai, China. This HPAI H5N1 virus outbreak affected large numbers of wild birds, such as bar-headed geese (*Anser indicus*), brown-headed gulls

Downloaded From Vet-eBooks.com

(*Larus brunnicephallus*), great black-headed gulls (*Larus ichthyaetus*), and great cormorants (*Phalacrocorax carbo*).[63,64] After the HPAI H5N1 virus outbreak in wild birds in 2005, the virus rapidly spread westward across Asia, Europe, the Middle East, and Africa, due to the movement of poultry and poultry products or via wild migratory birds.[60,65,66] The role of wild migratory birds in the spread of HPAI H5N1 is contentious as infected birds may be too severely affected to continue migration.[67] However, it has been shown that the pathogenesis of the HPAI H5N1 virus infection and the susceptibility of wild bird species to this infection varies considerably, depending on the bird species and previous exposure to influenza viruses. Experimental infections suggest that preexposure to LPAI viruses of homologous or heterologous subtypes may result in partial immunity to HPAI H5N1 virus infection.[39] Such preexisting immunity might not prevent viral replication but could protect birds from developing severe disease, thereby enabling them to continue to migrate and potentially spread the virus to other birds across large geographical areas. Upon experimental HPAI H5N1 virus infection, some duck species were found to develop either minor or no disease signs while still excreting the virus, predominantly from the respiratory tract, whereas other species developed a largely fatal infection that would not allow them to spread the virus efficiently over a considerable distance.[68–71] The outcome of HPAI H5N1 virus infections in wild bird species ranges from high morbidity and mortality (geese, swan, and certain duck species) to minimal morbidity without mortality (dabbling duck species). Therefore, although the spread of HPAI H5N1 into several parts of Asia was likely due to movement of poultry and poultry products,[60,65,66] the introductions into Europe were probably caused by migratory birds,[72,73] particularly as the affected regions had not reported outbreaks in poultry.[74,75] Although swan deaths have been the first indicator for the presence of the HPAI H5N1 virus in several European outbreaks, this does not necessarily implicate this species as primary vectors, but instead they could have been sentinel birds infected by other migrating bird species.

Although HPAI H5N1 viruses have gradually spread throughout regions of Asia, Europe, the Middle East, and Africa during the 2000s, the most dramatic and rapid emergence of HPAI viruses has been observed in early 2010s, when reassortant HPAI H5N8 viruses spread from Asia, into Europe and North America in less than 12 months. The H5N8 reassortant influenza viruses contained an HA that was highly similar to that of HPAI H5N1 viruses (from subclade 2.3.4.4), and began causing outbreaks in poultry in South Korea in early 2014.[76] The virus was thought to have been introduced into the summer breeding grounds of Beringia, a site that represents the convergence of a number of wild bird migratory flight paths. Coinciding with the autumn bird migration from Russia, the virus then spread west into Europe, resulting in poultry outbreaks in the Netherlands, Germany, the United Kingdom, and Italy, and east into North America.[77] The latter presented the first HPAI outbreak in North American poultry due to a Eurasian influenza virus.

Within North America, the HPAI H5N8 virus further reassorted with local LPAI viruses resulting in a novel HPAI H5N2 reassortant that contained five RNA segments from the H5N8 and three from North American LPAI viruses.[78] The H5N2 virus subsequently caused a substantial number of poultry outbreaks throughout the midwestern

Downloaded From Vet-eBooks.com

region of the United States in 2015. In addition, H5N1 viruses containing four RNA segments from the H5N8 virus and four from North American LPAI viruses were also detected in an apparently healthy duck in the United States during 2015.[79] The reasons why the reassortant HPAI H5NX viruses have emerged and spread so rapidly compared to H5N1 is unclear, but may relate to a difference in the pathogenicity caused by the viruses in key migratory bird species.[80] Besides the H5N6 virus in China, none of the other novel clade 2.3.4.4 H5NX viruses has caused human infections.

3. Evolutionary Genetics of Avian Influenza Viruses

3.1 Ecological Insights From Evolutionary Analysis: Natural Reservoirs

Geographic separation of avian host species has led to the segregation of contemporary LPAI viruses into two main phylogenetic lineages: the Eurasian and American lineages.[5,25] An example of the geographical differences between the genes within an HA subtype (H10 in this example) is given in Fig. 27.3, but similar phylogenies can be observed for other viral genes and subtypes. Despite these phylogenetic splits, the separation of the American and Eurasian wild bird and virus populations is not absolute. Some ducks and shorebirds cross the Bering Strait during migration or have breeding ranges that include both the Russian Far East and northwestern America. The majority of tundra shorebirds from the Russian Far East winter in Southeast Asia and Australia, but some species winter along the West coast of the Americas. The overlap in distribution of ducks is not as profound as that of shorebirds, but a few species (e.g., northern pintail, *Anas acuta*) are common in both North America and Eurasia and could also provide an intercontinental bridge for influenza A virus.[81] As a result, LPAI viruses carrying a mix of genes from the American and Eurasian lineages have been isolated occasionally, indicating that gene segments may be exchanged between the two virus populations.[22,82–87] Whole-genome analyses of LPAI virus isolates obtained from northern pintails in Alaska, a species that migrates between North America and Asia, suggested that intercontinental virus exchange can occur at a relatively high frequency.[85] Of the viruses analyzed in this study, a large proportion had at least one gene segment originating from the Eurasian lineage. In addition, analyses of H6 LPAI viruses pointed to introductions of the Eurasian H6 gene segment in North America on several occasions.[88] Of note is the establishment of the H10 subtype viruses in Australia, which contained the HA segment derived from the North American lineages. Phylogenetic analysis showed that mixing of the two gene pools occurred in migratory waterfowl prior to their introduction to Australia.[47] Although segments derived from cross-hemisphere transmission have been sporadically detected, virus strains with all eight gene segments derived from the other gene pool have not been detected.[82] An H11N2 LPAI virus was isolated from penguins in Antarctica, which was genetically distinct from all known contemporary influenza viruses, suggesting spatial separation from other lineages. The

Downloaded From Vet-eBooks.com

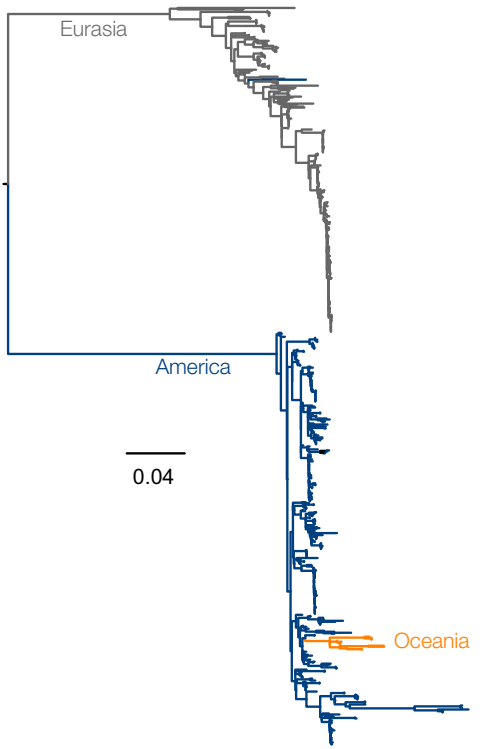


Figure 27.3 DNA maximum likelihood tree for the HA gene of subtype H10 influenza A viruses. All nearly full-length H10 HA genes available from GenBank were used to construct the tree. The major division into the Eurasian and American lineages of influenza viruses is clearly visible. The recent introduction and establishment of H10 viruses with a North American-lineage HA in Oceania (Australia) is further annotated.

geographic isolation of influenza virus hosts seems to be sufficient to facilitate the global circulation of contemporaneous LPAI in two separate gene pools. However, the genetic diversity of contemporaneous LPAI coalesced into a single lineage during the late 1800s, suggesting that the internal gene segments have undergone a "synchronized global sweep" beginning sometime in the late 1800s.[89]

Besides the influence of geographical separation on the evolutionary genetics of LPAI viruses, differences in host species have also resulted in clearly distinguishable virus populations. Good examples are the LPAI viruses of the H13 and H16 subtypes that are predominantly isolated from gulls and terns.[13,32] These viruses belong to a group of distinct LPAI viruses based on genetic, functional, and ecological properties and have evolved into separate genetic lineages from the viruses isolated from other *Charadriiformes* and *Anseriformes* (H1−H12 subtypes). Gene segments of gull viruses are genetically distinct from those circulating in other wild birds, suggesting that they have been separated for a sufficient amount of time to allow genetic differentiation by sympatric speciation.[13,25] An example of the diversification of the

Downloaded From Vet-eBooks.com

gull-lineage from other lineages of avian influenza virus genes is shown for gene segment 5 (NP) in Fig. 27.4.[90] Gull influenza viruses do not readily infect ducks upon experimental inoculation,[32] providing a biological explanation for the limited detection of these viruses in other avian influenza host species, although a limited number of gull viruses has been isolated from ducks and vice versa.[23,24,29] Genetic data from duck and shorebird influenza A virus isolates from the Americas suggests active transmission between these host populations, as genetic analyses did not reveal striking differences between viruses from these two groups of birds. This, therefore,

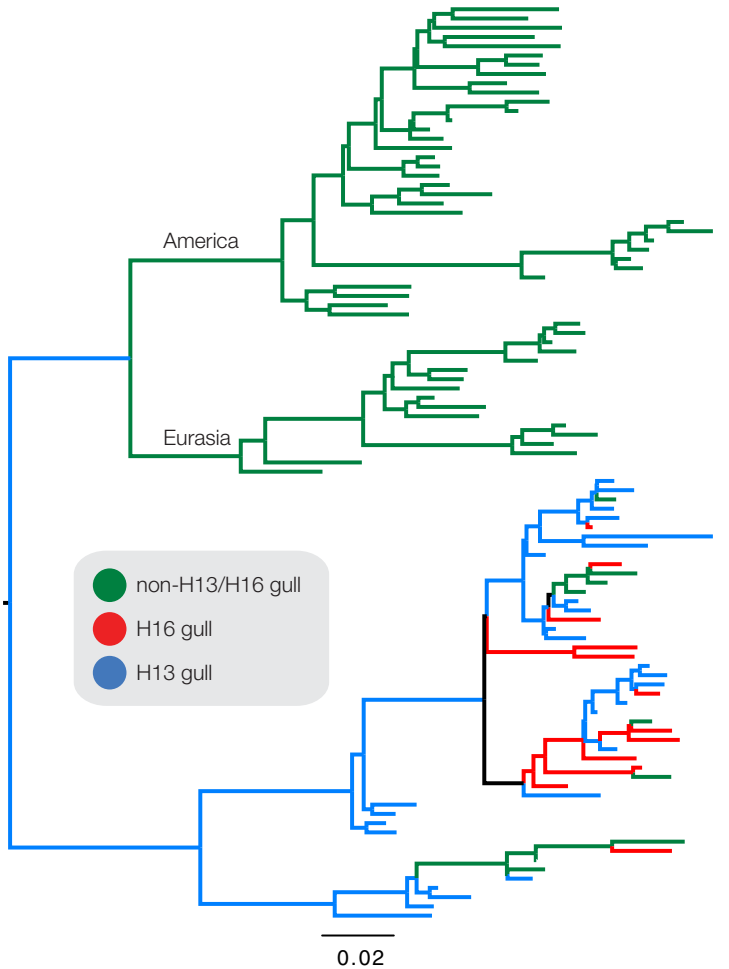


Figure 27.4 DNA maximum likelihood tree for the NP gene of gull influenza A viruses. All nearly full-length gull virus NP genes from GenBank were used to construct the tree. Virus of H13 and H16 subtypes form a genetic lineage (bottom), which is distinct from non-H13/H16 viruses detected in gulls. A separation is observed for American and Eurasian strains in both lineages.

Downloaded From Vet-eBooks.com

indicates that the duck and shorebird populations might function as one influenza virus host population.[91,92] Although certain HA subtypes have been reported to be more prevalent in either shorebirds or ducks in North America, this also does not seem to have resulted in differences in the genetic composition of influenza viruses obtained from these two reservoirs,[22,93] in contrast to what is observed for gulls.

The evolution of influenza viruses in natural hosts was considered to be in an "evolutionary stasis" suggesting that the genes evolved at a slower rate than in other hosts,[5] however, estimates of the rate of nucleotide substitution[82,94] have showed similar evolutionary rates to those seen in other hosts. For each gene segment, and within both the Eurasian and the American genetic lineages, multiple sublineages of viral genes seem to cocirculate, but without apparently consistent temporal or spatial correlations. The only additional noticeable peculiarity is the evolution of gene segment 8 (NS) into two highly divergent genetic lineages, referred to as alleles A and B (Fig. 27.5).[82,91] The biological significance of these two alleles remains unknown, but the alleles have been detected on both hemispheres, in all virus subtypes irrespective of wild bird host species.

The segmented genome of influenza viruses enables evolution by a process known as genetic reassortment, that is, the mixing of genes from two (or more) viruses.[5] Reassortment is one of the driving forces of the genetic variation of LPAI and HPAI viruses and contributes greatly to their phenotypic variability. A study of LPAI viruses obtained from Canadian ducks showed that genetic "sublineages" do not persist in wild birds, but frequently reassort.[95] Analysis of the genome constellation of five H4N6 LPAI viruses isolated from mallards on one day and one location revealed four different genome constellations with only one pair of viruses having an identical

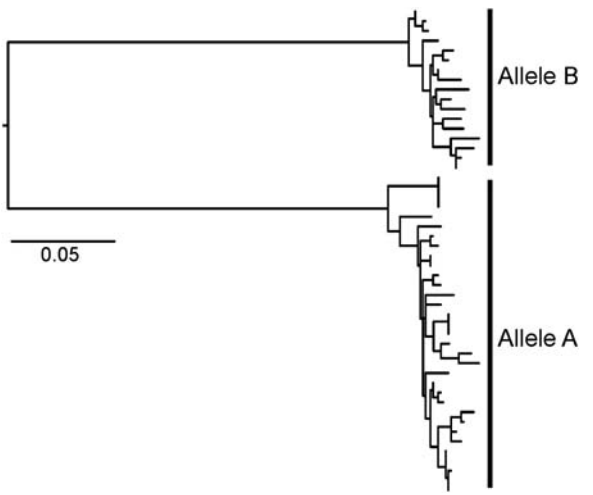


Figure 27.5 DNA maximum likelihood tree for the NS gene of avian influenza A viruses. All nearly full-length NS genes of Dutch avian viruses available from GenBank were used to construct the tree. These sequences are phylogenetically divided in two groups, representing two different "alleles," with no obvious correlation with time, location, or host species.

Downloaded From Vet-eBooks.com

genome composition.[82] The high LPAI virus prevalence in some wild bird species and the detection of concomitant infections in single birds support the notion that reassortment occurs at a relatively high rate in wild birds.[82,96] Thus, LPAI viruses do not present as "fixed" genome constellations, but reassortment leads to new "transient" genome constellations continuously.

3.2 Highly Pathogenic Avian Influenza H5N1 Virus

The evolution of the Eurasian HPAI H5N1 viruses since their first detection in 1997 has been well recorded.[97] Unique for HPAI viruses, the evolutionary path of H5N1 virus has been characterized by very frequent reassortment events, creating numerous novel so-called "genotypes" in time.[97] Thus, starting with viruses similar to A/Goose/Guangdong/1/1996, a series of reassortment events between HPAI H5N1 viruses and LPAI viruses from wild birds or poultry has led to the diverse H5N1 virus lineages circulating today. The wide range of domestic birds infected, and the contact of these birds with wild bird species may not only have provided opportunities for rapid spread of the virus to new areas, but also may have caused multiple cycles of periods of positive selection in new host species. Indeed, as compared to LPAI viruses in wild birds, the Asian H5N1 HPAI virus lineage displays a relatively high evolutionary rate and selection pressures, as measured by dN/dS ratios.[94]

As a consequence, the HA gene of the Eurasian HPAI H5N1 virus lineage has also evolved rapidly, diverging into a large and increasing number of antigenically and genetically distinct "clades"[98] (Fig. 27.6). Three new clade designations were recommended in 2015 based on division of clade 2.1.3.2a viruses in Indonesia, clade 2.2.1 viruses in Egypt, and clade 2.3.4 viruses that have been detected in Asia, Europe, and North America. The latter clade of viruses includes the newly emergent HPAI reassortant virus subtypes H5N2, H5N3, H5N5, H5N6, and H5N8. This evolutionary pattern is unprecedented in the recent history of HPAI viruses, and is likely due to multiple factors, including the enormous geographical spread of the outbreak throughout the eastern hemisphere, the involvement of a large number of hosts—both numbers of individual birds and numbers of different species—and the long duration of the outbreak, now lasting at least 18 years. The result of the high virus diversification is that preparation of preemptive pandemic vaccines to protect humans against the range of different HPAI H5 viruses circulating is extremely challenging. This is because vaccination against one clade of H5 virus will not provide protection against other circulating clades of H5 viruses. In the interest of both animal and human health, continued monitoring of emerging virus variants during new outbreaks is warranted.

4. Future Perspective

Wild bird surveillance programs have been implemented at an unprecedented scale in many parts of the world to determine the role of wild birds in the spread of avian influenza viruses and potentially to serve as a warning system for HPAI H5 virus incursions

Downloaded From Vet-eBooks.com

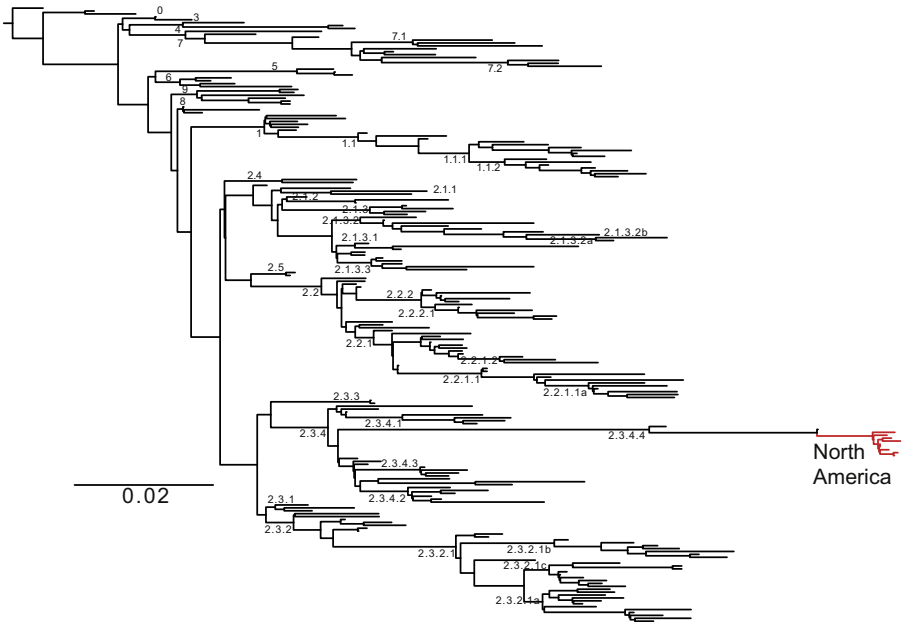


Figure 27.6 DNA maximum likelihood tree for the HA gene of representative HPAI H5N1 viruses. Nearly full-length HA genes available from GenBank were used to construct the tree. The phylogenetic tree shows the known "clades" of viruses, as identified by numbers. These clades were defined using criteria based on sharing of a common ancestor and monophyletic grouping with high bootstrap support. The tree also illustrates the phylogenetic position of the newly emerged H5NX reassortant viruses detected in North America during 2014–2015.

into new geographical regions. At the same time, the increased capacity of current day sequencing technology has facilitated the genetic analyses of large numbers of influenza A virus genomes. Together, these developments provide a unique opportunity to advance our understanding of the ecology and evolution of avian influenza in wild birds, and in the longer term will help to limit the impact of influenza on human and animal health. To this end, we hope that all specialists in the field—including virologists, epidemiologists, ornithologists, geneticists, ecologists, veterinarians, clinicians, mathematicians, and bioinformaticists—will work together closely to make optimal use of the wealth of data available, ultimately leading to a better understanding of the ecology of avian influenza—LPAI and HPAI—in wild birds.

Acknowledgments

We acknowledge Josanne Verhagen and Vincent Munster who contributed to the "Ecology and evolution of avian influenza viruses" chapter in the 1st edition of *Genetics and Evolution of Infectious Diseases*, which was the basis of this updated version. The Melbourne WHO Collaborating Centre for Reference and Research on Influenza is supported by the Australian Government Department of Health.

Downloaded From Vet-eBooks.com

References

1. Palese PS, Shaw ML. Orthomyxoviridae: the viruses and their replication. In: Knipe DMH PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Strauss S, editors. *Fields virology*. 5th ed. Philadelphia: Lippincott Williams and Wilkins; 2007. p. 1647−90.
2. Wright PF, Neumann G, Kawaoka Y. Orthomyxoviruses. In: Knipe DMH PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Strauss S, editors. *Fields virology*. 5th ed. Philadelphia: Lippincott Williams and Wilkins; 2007. p. 1647−90.
3. Ducatez MF, Pelletier C, Meyer G. Influenza D virus in cattle, France, 2011−2014. *Emerg Infect Dis* 2015;**21**(2):368−71.
4. ICTV. *Virus taxonomy, release of the international committee on taxonomy of viruses*. 2014. Available from: www.ictvonline.org/virusTaxonomy.asp.
5. Yoon SW, Webby RJ, Webster RG. Evolution and ecology of influenza A viruses. *Curr Top Microbiol Immunol* 2014;**385**:359−75.
6. Chen W, Calvo PA, Malide D, Gibbs J, Schubert U, Bacik I, et al. A novel influenza A virus mitochondrial protein that induces cell death. *Nat Med* 2001;**7**(12):1306−12.
7. Conenello GM, Palese P. Influenza A virus PB1-F2: a small protein with a big punch. *Cell Host Microbe* 2007;**2**(4):207−9.
8. Wise HM, Foeglein A, Sun J, Dalton RM, Patel S, Howard W, et al. A complicated message: identification of a novel PB1-related protein translated from influenza A virus segment 2 mRNA. *J Virol* 2009;**83**(16):8021−31.
9. Jagger BW, Wise HM, Kash JC, Walters KA, Wills NM, Xiao YL, et al. An overlapping protein-coding region in influenza A virus segment 3 modulates the host response. *Science* 2012;**337**(6091):199−204.
10. Tong S, Li Y, Rivailler P, Conrardy C, Castillo DA, Chen LM, et al. A distinct lineage of influenza A virus from bats. *Proc Natl Acad Sci USA* 2012;**109**(11):4269−74.
11. Tong S, Zhu X, Li Y, Shi M, Zhang J, Bourgeois M, et al. New world bats harbor diverse influenza A viruses. *PLoS Pathog* 2013;**9**(10):e1003657.
12. Garcia-Sastre A. The neuraminidase of bat influenza viruses is not a neuraminidase. *Proc Natl Acad Sci USA* 2012;**109**(46):18635−6.
13. Fouchier RA, Munster V, Wallensten A, Bestebroer TM, Herfst S, Smith D, et al. Characterization of a novel influenza A virus hemagglutinin subtype (H16) obtained from black-headed gulls. *J Virol* 2005;**79**(5):2814−22.
14. Karlsson Hedestam GB, Fouchier RA, Phogat S, Burton DR, Sodroski J, Wyatt RT. The challenges of eliciting neutralizing antibodies to HIV-1 and to influenza virus. *Nat Rev Microbiol* 2008;**6**(2):143−55.
15. Alexander DJ. A review of avian influenza in different bird species. *Vet Microbiol* 2000; **74**(1−2):3−13.
16. Alexander DJ. An overview of the epidemiology of avian influenza. *Vaccine* 2007;**25**(30): 5637−44.
17. Webster RG, Rott R. Influenza virus A pathogenicity: the pivotal role of hemagglutinin. *Cell* 1987;**50**(5):665−6.
18. Ellstrom P, Latorre-Margalef N, Griekspoor P, Waldenstrom J, Olofsson J, Wahlgren J, et al. Sampling for low-pathogenic avian influenza A virus in wild Mallard ducks: oropharyngeal versus cloacal swabbing. *Vaccine* 2008;**26**(35):4414−6.
19. Hanson BA, Stallknecht DE, Swayne DE, Lewis LA, Senne DA. Avian influenza viruses in Minnesota ducks during 1998−2000. *Avian Dis* 2003;**47**(Suppl. 3):867−71.

Downloaded From Vet-eBooks.com

20. Hinshaw VS, Webster RG, Bean WJ, Sriram G. The ecology of influenza viruses in ducks and analysis of influenza viruses with monoclonal antibodies. *Comp Immunol Microbiol Infect Dis* 1980;**3**(1−2):155−64.
21. Hinshaw VS, Webster RG, Turner B. The perpetuation of orthomyxoviruses and paramyxoviruses in Canadian waterfowl. *Can J Microbiol* 1980;**26**(5):622−9.
22. Krauss S, Obert CA, Franks J, Walker D, Jones K, Seiler P, et al. Influenza in migratory birds and evidence of limited intercontinental virus exchange. *PLoS Pathog* 2007;**3**(11): e167.
23. Krauss S, Walker D, Pryor SP, Niles L, Chenghong L, Hinshaw VS, et al. Influenza A viruses of migrating wild aquatic birds in North America. *Vector Borne Zoonotic Dis* 2004; **4**(3):177−89.
24. Munster VJ, Baas C, Lexmond P, Waldenstrom J, Wallensten A, Fransson T, et al. Spatial, temporal, and species variation in prevalence of influenza A viruses in wild migratory birds. *PLoS Pathog* 2007;**3**(5):e61.
25. Olsen B, Munster VJ, Wallensten A, Waldenstrom J, Osterhaus AD, Fouchier RA. Global patterns of influenza A virus in wild birds. *Science* 2006;**312**(5772):384−8.
26. Parmley EJ, Bastien N, Booth TF, Bowes V, Buck PA, Breault A, et al. Wild bird influenza survey, Canada, 2005. *Emerg Infect Dis* 2008;**14**(1):84−7.
27. Wallensten A, Munster VJ, Karlsson M, Lundkvist A, Brytting M, Stervander M, et al. High prevalence of influenza A virus in ducks caught during spring migration through Sweden. *Vaccine* 2006;**24**(44−46):6734−5.
28. Wallensten A, Munster VJ, Latorre-Margalef N, Brytting M, Elmberg J, Fouchier RA, et al. Surveillance of influenza A virus in migratory waterfowl in northern Europe. *Emerg Infect Dis* 2007;**13**(3):404−11.
29. Kawaoka Y, Chambers TM, Sladen WL, Webster RG. Is the gene pool of influenza viruses in shorebirds and gulls different from that in wild ducks? *Virology* 1988;**163**(1):247−50.
30. Gaidet N, Newman SH, Hagemeijer W, Dodman T, Cappelle J, Hammoumi S, et al. Duck migration and past influenza A (H5N1) outbreak areas. *Emerg Infect Dis* 2008;**14**(7): 1164−6.
31. Hurt AC, Hansbro PM, Selleck P, Olsen B, Minton C, Hampson AW, et al. Isolation of avian influenza viruses from two different transhemispheric migratory shorebird species in Australia. *Arch Virol* 2006;**151**(11):2301−9.
32. Hinshaw VS, Air GM, Gibbs AJ, Graves L, Prescott B, Karunakaran D. Antigenic and genetic characterization of a novel hemagglutinin subtype of influenza A viruses from gulls. *J Virol* 1982;**42**(3):865−72.
33. Pasick J, Berhane Y, Kehler H, Hisanaga T, Handel K, Robinson J, et al. Survey of influenza A viruses circulating in wild birds in Canada 2005 to 2007. *Avian Dis* 2010;**54**(Suppl. 1): 440−5.
34. Grillo VL, Arzey KE, Hansbro PM, Hurt AC, Warner S, Bergfeld J, et al. Avian influenza in Australia: a summary of 5 years of wild bird surveillance. *Aust Vet J* 2015;**93**(11):387−93.
35. Muzyka D, Pantin-Jackwood M, Starick E, Fereidouni S. Evidence for genetic variation of Eurasian avian influenza viruses of subtype H15: the first report of an H15N7 virus. *Arch Virol* 2016;**161**(3):605−12.
36. Bowman AS, Nolting JM, Massengill R, Baker J, Workman JD, Slemons RD. Influenza A virus surveillance in waterfowl in Missouri, USA, 2005−2013. *Avian Dis* 2015;**59**(2): 303−8.
37. Halvorson DA, Kelleher CJ, Senne DA. Epizootiology of avian influenza: effect of season on incidence in sentinel ducks and domestic turkeys in Minnesota. *Appl Environ Microbiol* 1985;**49**(4):914−9.

Downloaded From Vet-eBooks.com

38. Hinshaw VS, Webster RG, Turner B. Novel influenza A viruses isolated from Canadian feral ducks: including strains antigenically related to swine influenza (Hsw1N1) viruses. *J Gen Virol* 1978;**41**(1):115–27.
39. Fereidouni SR, Starick E, Beer M, Wilking H, Kalthoff D, Grund C, et al. Highly pathogenic avian influenza virus infection of mallards with homo- and heterosubtypic immunity induced by low pathogenic avian influenza viruses. *PLoS One* 2009;**4**(8):e6706.
40. Jourdain E, Gunnarsson G, Wahlgren J, Latorre-Margalef N, Brojer C, Sahlin S, et al. Influenza virus in a natural host, the mallard: experimental infection data. *PLoS One* 2010; **5**(1):e8935.
41. Kida H, Yanagawa R, Matsuoka Y. Duck influenza lacking evidence of disease signs and immune response. *Infect Immun* 1980;**30**(2):547–53.
42. Latorre-Margalef N, Gunnarsson G, Munster VJ, Fouchier RA, Osterhaus AD, Elmberg J, et al. Effects of influenza A virus infection on migrating mallard ducks. *Proc Biol Sci R Soc* 2009;**276**(1659):1029–36.
43. Verhagen JH, Majoor F, Lexmond P, Vuong O, Kasemir G, Lutterop D, et al. Epidemiology of influenza A virus among black-headed gulls, the Netherlands, 2006–2010. *Emerg Infect Dis* 2014;**20**(1):138–41.
44. Hanson BA, Luttrell MP, Goekjian VH, Niles L, Swayne DE, Senne DA, et al. Is the occurrence of avian influenza virus in *Charadriiformes* species and location dependent? *J Wildl Dis* 2008;**44**(2):351–61.
45. Haynes L, Arzey E, Bell C, Buchanan N, Burgess G, Cronan V, et al. Australian surveillance for avian influenza viruses in wild birds between July 2005 and June 2007. *Aust Vet J* 2009;**87**(7):266–72.
46. Winker K, Spackman E, Swayne DE. Rarity of influenza A virus in spring shorebirds, southern Alaska. *Emerg Infect Dis* 2008;**14**(8):1314–6.
47. Vijaykrishna D, Deng YM, Su YC, Fourment M, Iannello P, Arzey GG, et al. The recent establishment of North American H10 lineage influenza viruses in Australian wild waterfowl and the evolution of Australian avian influenza viruses. *J Virol* 2013;**87**(18):10182–9.
48. Ferenczi M, Beckmann C, Warner S, Loyn R, O'Riley K, Wang X, et al. Avian influenza infection dynamics under variable climatic conditions, viral prevalence is rainfall driven in waterfowl from temperate, south-east Australia. *Veterinary Res* 2016;**47**(1):23.
49. Kleijn D, Munster VJ, Ebbinge BS, Jonkers DA, Muskens GJ, Van Randen Y, et al. Dynamics and ecological consequences of avian influenza virus infection in greater white-fronted geese in their winter staging areas. *Proc Biol Sci R Soc* 2010;**277**(1690):2041–8.
50. Brown JD, Goekjian G, Poulson R, Valeika S, Stallknecht DE. Avian influenza virus in water: infectivity is dependent on pH, salinity and temperature. *Vet Microbiol* 2009; **136**(1–2):20–6.
51. Stallknecht DE, Brown JD. Tenacity of avian influenza viruses. *Revue Sci Tech* 2009;**28**(1): 59–67.
52. Stallknecht DE, Kearney MT, Shane SM, Zwank PJ. Effects of pH, temperature, and salinity on persistence of avian influenza viruses in water. *Avian Dis* 1990;**34**(2):412–8.
53. Stallknecht DE, Shane SM, Kearney MT, Zwank PJ. Persistence of avian influenza viruses in water. *Avian Dis* 1990;**34**(2):406–11.
54. WHO. *WHO risk assessment of human infections with avian influenza A(H7N9) virus*. February 23, 2015. Available from: http://www.who.int/influenza/human_animal_interface/influenza_h7n9/RiskAssessment_H7N9_23Feb20115.pdf.
55. OIE. *Update on highly pathogenic avian influenza in animals (type H5 and H7)*. 2015. Available from: http://www.oie.int/en/animal-health-in-the-world/update-on-avian-influenza/2015/.

Downloaded From Vet-eBooks.com

56. Alexander DJ, Brown IH. History of highly pathogenic avian influenza. *Revue Sci Tech* 2009;**28**(1):19−38.
57. Chen H, Deng G, Li Z, Tian G, Li Y, Jiao P, et al. The evolution of H5N1 influenza viruses in ducks in southern China. *Proc Natl Acad Sci USA* July 13, 2004;**101**(28):10452−7.
58. de Jong JC, Claas EC, Osterhaus AD, Webster RG, Lim WL. A pandemic warning? *Nature* October 9, 1997;**389**(6651):554. PubMed PMID: 9335492.
59. Ellis TM, Bousfield RB, Bissett LA, Dyrting KC, Luk GS, Tsim ST, et al. Investigation of outbreaks of highly pathogenic H5N1 avian influenza in waterfowl and wild birds in Hong Kong in late 2002. Avian pathology. *J WVPA* 2004;**33**(5):492−505. PubMed PMID: 15545029.
60. Sims LD, Ellis TM, Liu KK, Dyrting K, Wong H, Peiris M, et al. Avian influenza in Hong Kong 1997−2002. *Avian Dis* 2003;**47**(Suppl. 3):832−8.
61. Sturm-Ramirez KM, Ellis T, Bousfield B, Bissett L, Dyrting K, Rehg JE, et al. Reemerging H5N1 influenza viruses in Hong Kong in 2002 are highly pathogenic to ducks. *J Virol* 2004; **78**(9):4892−901.
62. Vijaykrishna D, Bahl J, Riley S, Duan L, Zhang JX, Chen H, et al. Evolutionary dynamics and emergence of panzootic H5N1 influenza viruses. *PLoS Pathog* 2008;**4**(9):e1000161.
63. Chen H, Smith GJ, Zhang SY, Qin K, Wang J, Li KS, et al. Avian flu: H5N1 virus outbreak in migratory waterfowl. *Nature* 2005;**436**(7048):191−2.
64. Liu J, Xiao H, Lei F, Zhu Q, Qin K, Zhang XW, et al. Highly pathogenic H5N1 influenza virus infection in migratory birds. *Science* 2005;**309**(5738):1206.
65. Alexander DJ, Brown IH. Recent zoonoses caused by influenza A viruses. *Revue Sci Tech* April 2000;**19**(1):197−225.
66. Gilbert M, Chaitaweesub P, Parakamawongsa T, Premashthira S, Tiensin T, Kalpravidh W, et al. Free-grazing ducks and highly pathogenic avian influenza, Thailand. *Emerg Infect Dis* 2006;**12**(2):227−34.
67. Feare CJ, Yasue M. Asymptomatic infection with highly pathogenic avian influenza H5N1 in wild birds: how sound is the evidence? *Virol J* 2006;**3**:96.
68. Brown JD, Stallknecht DE, Beck JR, Suarez DL, Swayne DE. Susceptibility of North American ducks and gulls to H5N1 highly pathogenic avian influenza viruses. *Emerg Infect Dis* 2006;**12**(11):1663−70.
69. Brown JD, Stallknecht DE, Swayne DE. Experimental infection of swans and geese with highly pathogenic avian influenza virus (H5N1) of Asian lineage. *Emerg Infect Dis* January 2008;**14**(1):136−42.
70. Kalthoff D, Breithaupt A, Teifke JP, Globig A, Harder T, Mettenleiter TC, et al. Highly pathogenic avian influenza virus (H5N1) in experimentally infected adult mute swans. *Emerg Infect Dis* 2008;**14**(8):1267−70.
71. Keawcharoen J, van Riel D, van Amerongen G, Bestebroer T, Beyer WE, van Lavieren R, et al. Wild ducks as long-distance vectors of highly pathogenic avian influenza virus (H5N1). *Emerg Infect Dis* 2008;**14**(4):600−7.
72. Si Y, Skidmore AK, Wang T, de Boer WF, Debba P, Toxopeus AG, et al. Spatio-temporal dynamics of global H5N1 outbreaks match bird migration patterns. *Geospatial Health* 2009;**4**(1):65−78.
73. Starick E, Beer M, Hoffmann B, Staubach C, Werner O, Globig A, et al. Phylogenetic analyses of highly pathogenic avian influenza virus isolates from Germany in 2006 and 2007 suggest at least three separate introductions of H5N1 virus. *Vet Microbiol* 2008; **128**(3−4):243−52.
74. Globig A, Baumer A, Revilla-Fernandez S, Beer M, Wodak E, Fink M, et al. Ducks as sentinels for avian influenza in wild birds. *Emerg Infect Dis* 2009;**15**(10):1633−6.

Downloaded From Vet-eBooks.com

75. Hesterberg U, Harris K, Stroud D, Guberti V, Busani L, Pittman M, et al. Avian influenza surveillance in wild birds in the European Union in 2006. *Influenza Other Respir Viruses* 2009;**3**(1):1−14.
76. Lee YJ, Kang HM, Lee EK, Song BM, Jeong J, Kwon YK, et al. Novel reassortant influenza A(H5N8) viruses, South Korea. *Emerg Infect Dis* 2014;**20**(6):1087−9.
77. Lee DH, Torchetti MK, Winker K, Ip HS, Song CS, Swayne DE. Intercontinental spread of Asian-origin H5N8 to North America through Beringia by migratory birds. *J Virol* 2015; **89**(12):6521−4.
78. Pasick J, Berhane Y, Joseph T, Bowes V, Hisanaga T, Handel K, et al. Reassortant highly pathogenic influenza A H5N2 virus containing gene segments related to Eurasian H5N8 in British Columbia, Canada, 2014. *Sci Rep* 2015;**5**:9484.
79. Torchetti MK, Killian ML, Dusek RJ, Pedersen JC, Hines N, Bodenstein B, et al. Novel H5 clade 2.3.4.4 reassortant (H5N1) virus from a green-winged teal in Washington, USA. *Genome Announc* 2015;**3**(2).
80. Kang HM, Lee EK, Song BM, Jeong J, Choi JG, Jeong J, et al. Novel reassortant influenza A(H5N8) viruses among inoculated domestic and wild ducks, South Korea, 2014. *Emerg Infect Dis* 2015;**21**(2):298−304.
81. Verhagen JH, van der Jeugd HP, Nolet BA, Slaterus R, Kharitonov SP, de Vries PP, et al. Wild bird surveillance around outbreaks of highly pathogenic avian influenza A(H5N8) virus in the Netherlands, 2014, within the context of global flyways. *Eurosurveillance: Bulletin Europeen sur les maladies transmissibles [Eur Commun Dis Bull]* 2015;**20**(12).
82. Dugan VG, Chen R, Spiro DJ, Sengamalay N, Zaborsky J, Ghedin E, et al. The evolutionary genetics and emergence of avian influenza viruses in wild birds. *PLoS Pathog* 2008;**4**(5): e1000076. PubMed PMID: 18516303.
83. Koehler AV, Pearce JM, Flint PL, Franson JC, Ip HS. Genetic evidence of intercontinental movement of avian influenza in a migratory bird: the northern pintail (*Anas acuta*). *Mol Ecol* November 2008;**17**(21):4754−62.
84. Makarova NV, Kaverin NV, Krauss S, Senne D, Webster RG. Transmission of Eurasian avian H2 influenza virus to shorebirds in North America. *J Gen Virol* 1999;**80**(Pt. 12):3167−71.
85. Ramey AM, Pearce JM, Flint PL, Ip HS, Derksen DV, Franson JC, et al. Intercontinental reassortment and genomic variation of low pathogenic avian influenza viruses isolated from northern pintails (*Anas acuta*) in Alaska: examining the evidence through space and time. *Virology* 2010;**401**(2):179−89.
86. Wahlgren J, Waldenstrom J, Sahlin S, Haemig PD, Fouchier RA, Osterhaus AD, et al. Gene segment reassortment between American and Asian lineages of avian influenza virus from waterfowl in the Beringia area. *Vector Borne Zoonotic Dis* 2008;**8**(6):783−90.
87. Wallensten A, Munster VJ, Elmberg J, Osterhaus AD, Fouchier RA, Olsen B. Multiple gene segment reassortment between Eurasian and American lineages of influenza A virus (H6N2) in Guillemot (*Uria aalge*). *Arch Virol* 2005;**150**(8):1685−92.
88. zu Dohna H, Li J, Cardona CJ, Miller J, Carpenter TE. Invasions by Eurasian avian influenza virus H6 genes and replacement of the virus' North American clade. *Emerg Infect Dis* 2009;**15**(7):1040−5.
89. Worobey M, Han GZ, Rambaut A. A synchronized global sweep of the internal genes of modern avian influenza virus. *Nature* 2014;**508**(7495):254−7.
90. Gorman OT, Bean WJ, Kawaoka Y, Webster RG. Evolution of the nucleoprotein gene of influenza A virus. *J Virol* 1990;**64**(4):1487−97.
91. Spackman E, Stallknecht DE, Slemons RD, Winker K, Suarez DL, Scott M, et al. Phylogenetic analyses of type A influenza genes in natural reservoir species in North America reveals genetic variation. *Virus Res* 2005;**114**(1−2):89−100.

Downloaded From Vet-eBooks.com

92. Widjaja L, Krauss SL, Webby RJ, Xie T, Webster RG. Matrix gene of influenza a viruses isolated from wild aquatic birds: ecology and emergence of influenza a viruses. *J Virol* 2004;**78**(16):8771−9.
93. Obenauer JC, Denson J, Mehta PK, Su X, Mukatira S, Finkelstein DB, et al. Large-scale sequence analysis of avian influenza isolates. *Science* 2006;**311**(5767):1576−80.
94. Chen R, Holmes EC. Avian influenza virus exhibits rapid evolutionary dynamics. *Mol Biol Evol* 2006;**23**(12):2336−41.
95. Hatchette TF, Walker D, Johnson C, Baker A, Pryor SP, Webster RG. Influenza A viruses in feral Canadian ducks: extensive reassortment in nature. *J Gen Virol* 2004;**85**(Pt. 8): 2327−37.
96. Macken CA, Webby RJ, Bruno WJ. Genotype turnover by reassortment of replication complex genes from avian influenza A virus. *J Gen Virol* 2006;**87**(Pt. 10):2803−15.
97. Neumann G, Green MA, Macken CA. Evolution of highly pathogenic avian H5N1 influenza viruses and the emergence of dominant variants. *J Gen Virol* 2010;**91**(Pt. 8):1984−95.
98. Smith GJ, Donis RO. World Health Organization/World Organisation for Animal HF, Agriculture Organization HEWG. Nomenclature updates resulting from the evolution of avian influenza A(H5) virus clades 2.1.3.2a, 2.2.1, and 2.3.4 during 2013−2014. *Influenza Other Respir Viruses* 2015;**9**(5):271−6.

Downloaded From Vet-eBooks.com

Index

'*Note:* Page numbers followed by "f" indicate figures, "t" indicate tables and "b" indicates boxes.'

A
AAC(6′)-Ib-cr gene, 267
AAG. *See* Albumin and α1-acid glycoprotein (AAG)
AAT. *See* Animal trypanosomiasis (AAT)
Abacavir (ABC), 435−436
ABC methods. *See* Approximate Bayesian computation methods (ABC methods)
ABI. *See* Applied Biosystems Inc. (ABI)
Abiotic environment effect on coevolution, 123−124
Absorption, distribution, metabolism, and excretion pathways (ADME pathways), 435
Acari, 367−370
ACE2. *See* Angiotensin-converting enzyme 2 (ACE2)
Acetylcholine (ACh), 325
Acetylcholinesterase (AChE), 316, 325−326
N-Acetyltransferase 2 (NAT2), 445
ACh. *See* Acetylcholine (ACh)
AChE. *See* Acetylcholinesterase (AChE)
ACME. *See* Arginine catabolic mobile element (ACME)
Acquired immune deficiency syndrome (AIDS), 573, 574f, 579−580. *See also* Human immunodeficiency virus (HIV)
ACT. *See* Artemisinin-based combination therapy (ACT)
ACTG. *See* AIDS Clinical trials (ACTG)
Adapt Globally Act Locally model, 37
Adaptations, 314
  evolution of proteins encoding for antibiotic-resistance genes, 267−268
  immune system, 195
  introgression, 365
  limitations and cost of resistance
    beyond model organisms, 273−274
    genetics of adaptation, 269−271
    genotype to phenotype, 271−273
AdaptML, 35
ADLI. *See* Antituberculous drug-induced liver injury (ADLI)
ADME pathways. *See* Absorption, distribution, metabolism, and excretion pathways (ADME pathways)
AE1. *See* Ancestral Europe 1 (AE1)
Aedes aegypti (*A. aegypti*), 236, 286
Aedes albopictus (*A. albopictus*), 294−295, 313−314
Af1 subtype. *See* African 1 subtype (Af1 subtype)
Affymetrix database, 427
AFLP. *See* Amplified fragment length polymorphism (AFLP)
African 1 subtype (Af1 subtype), 537
African 2 subtype (Af2 subtype), 537
African apes
  HIV-1 from SIV among, 581−585
  infections in Old World monkeys and, 576t−577t
African non-human primates, SIV in, 575−579
African trypanosomiasis, genetics of, 343−351
AgdbNet software, 388
Agriculture, 198−199
AIC. *See* Akaike Information Criterion (AIC)
AIDS. *See* Acquired immune deficiency syndrome (AIDS)
AIDS Clinical trials (ACTG), 438−439
Akaike Information Criterion (AIC), 392
AL. *See* Artemether−lumefantrine (AL)

Downloaded From Vet-eBooks.com

Albumin and α1-acid glycoprotein (AAG), 439–440
Allele-based methods, 390–392
Alleles A and B, 632
Allelic profiles database, 388
Allometry, 289
Alternative insecticides, 317–318
American trypanosomiasis. *See* Chagas disease
γ-Aminobutyric acid receptor (GABA receptor), 315, 322–323
Aminoglycoside acetyltransferase, 267
Amodiaquine (AQ), 443, 498
AMOVA framework, 84
AmpC proteins, 264–265
Amplified fragment length polymorphism (AFLP), 172
Analytical approaches, 390–394
  allele-based methods, 391–392
  nucleotide-based methods, 392–394
Anas penelope. *See* Eurasian wigeons (*Anas penelope*)
Anatomical landmarks, 286
Ancestral Europe 1 (AE1), 145
Ancestral Europe 2 (AE2), 145
Ancestral HPAI H5N1 virus, 627
Angiotensin-converting enzyme 2 (ACE2), 612–613
Animal and plant pathogens comparison, 87–88
Animal link, 605–606
Animal trypanosomiasis (AAT), 343
Animal-adapted strains, 538
Anopheles coluzzii (*A. coluzzii*), 365
Anopheles gambiae (*A. gambiae*), 212, 361–362
  complex, 361–367
  conclusions, 367
  geographical distribution of species, 362f
  population genetic structure, 363–365
  population genetics to malaria transmission and control, 366
  relationship between *A. coluzzii* and, 365
Anopheles Plasmodium-responsive leucine-rich repeat 1 (APL1), 217
Anopheles stephensi (*A. stephensi*), 317–318
Anseriformes, 621, 624, 630–632
Antagonistic coevolution, 116–119, 121
  effect of abiotic environment, 123–124
  environmental and community context, 122–123
  generation time, 122
  influence of microbiome on host–pathogen interactions, 123
  migration, mutation, and recombination, 121–122
Antagonistic pleiotropy, 267
Antibiotic resistance, 257. *See also* Insecticide resistance
  biological units, 258
  conventional scientific wisdom, 274–275
  determinants, 560
  evolution of antibiotic-resistance genes, 264–269
  gene evolution, 264–269
    adaptive evolution of proteins encoding for, 267–268
    β-lactamases model, 268–269
    as targets, 264–267
  limitations to adaptation and cost of resistance, 269–274
  mechanisms and sources, 260–264, 260f
  mutants, 270
Antimalarial therapy, pharmacogenetics of. *See also* HIV therapy, pharmacogenetics of
  amodiaquine, 443
  artemisinin compounds, 442–443
  mefloquine, 443–444
  primaquine, 443
  proguanil, 444
  quinine, 444
Antiretroviral treatment (ART), 573
Antituberculous drug-induced liver injury (ADLI), 445
Antituberculous therapy, pharmacogenetics of, 444–445
APL1. *See Anopheles Plasmodium*-responsive leucine-rich repeat 1 (APL1)
APOA5. *See* Apolipoprotein A5 (APOA5)
APOBEC3G, 589
APOL-1. *See* Apolipoprotein L1 (APOL-1)
Apolipoprotein A5 (APOA5), 441
Apolipoprotein L1 (APOL-1), 195–196
Applied Biosystems Inc. (ABI), 421

Downloaded From Vet-eBooks.com

Approximate Bayesian computation methods (ABC methods), 63
AQ. *See* Amodiaquine (AQ)
Arboviruses, 367
Archaea, 103
Arenaria interpres. *See* Ruddy turnstones (*Arenaria interpres*)
Arginine catabolic mobile element (ACME), 563–564
"Arms-race" process, 86–87, 116–119
ART. *See* Antiretroviral treatment (ART)
Artemether–lumefantrine (AL), 499–500
Artemisinin
  compounds, 442–443
  parasite resistance evolution in artemisinin combination therapy era, 499–500
Artemisinin-based combination therapy (ACT), 216, 442
Artesunate (AS), 442
Artesunate–amodiaquine (ASAQ), 499–500
Arthropods
  allometry, 289
  applications in medical entomology, 296–301
  CD, 293–294
  genetics of metric change, 290–293
  landmark-based geometric morphometry, 285–288
  measurement error, 289
  modern and traditional morphometrics, 285
  pseudo-landmark-based shape, 288
  regulation of phenotype, 294–295
Artifacts, 411
AS. *See* Artesunate (AS)
ASAQ. *See* Artesunate–amodiaquine (ASAQ)
Ascomycetes, 73
Ascomycota, 71–72
Asexuality
  asexual reproduction, 80, 99–100
  in biosphere, 103
  genetic consequences, 103–105
Asparagine, 613
Aspergillus fumigatus (*A. fumigatus*), 73
Aspidiotus perniciosus (*A. perniciosus*), 315
Associations and other genes, 514–515
Atazanavir (ATV), 435
ATV. *See* Atazanavir (ATV)
ATV/r. *See* ATV/ritonavir (ATV/r)
ATV/ritonavir (ATV/r), 435
Atypical pneumonia. *See* Severe Acute Respiratory Syndrome (SARS)
Augmented tree, 186
Avian influenza viruses
  evolutionary genetics, 629–633
  future perspective, 633–634
  influenza A virus, 621–623
  influenza viruses, 624–629

B

b2m. *See* Beta-2-microglobulin (b2m)
Babesiosis, 367
Bacillus anthracis (*B. anthracis*), 41
Bacillus mallei (*B. mallei*), 396–397
Bacillus pseudomallei (*B. pseudomallei*), 396–397
Bacillus sphaericus (Bs), 317
Bacillus thailandensis (*B. thailandensis*), 396–397
Bacillus thuringiensis (Bt), 41, 317
BACs. *See* Bacterial artificial chromosomes (BACs)
Bacteria(l), 103
  ecology into bacterial systematics, 43–44
  ecotype cohesion, 41–42
  epidemic clone, 263
  H. pylori, 144–150
  M. tuberculosis, 150–155
  pathogens, 37, 405–406
  recombination in bacterial populations
    emergence and persistence of sequence clusters, 52–54
    gene flow across species boundaries, 57
    heterogeneity in recombination, 54–55
    pan-genome of species and populations, 55–57
  speciation
    models, 27f
    Stable Ecotype model, 33–35
  species, 25, 383, 385–387, 396–397
  typing, 388
Bacterial artificial chromosomes (BACs), 530–531
Bacteriophages, 405–406
BAPS approach, 61
Based Upon Related Sequences Types approach (eBURST approach), 391

Downloaded From Vet-eBooks.com

Basic polymerase 1 (PB1), 621−623
Basic polymerase 2 (PB2), 621−623
Basidiomycetes, 73−74
Basidiomycota, 71−72
Bat SL-CoV genomic structure, 604f
Batrachochytrium dendrobatidis (*Bd*), 75−76
  global emergence of Amphibian Pathogen, 75−77
Bayesian
  clustering analyses, 145, 150−151
  clustering method, 143, 147−149
  coalescent approach, 393
  MCMC approach, 157−158
  model, 62
  skyline plots, 158f
Bayesian Evolutionary Analysis Sampling Trees (BEAST), 62
Bayesian inference (BI), 392−393
Bayesian Information Criterion (BIC), 392
BCG vaccine, 541
BCRP/ABCG2. *See* Breast cancer-related protein (BCRP/ABCG2)
Bd. *See Batrachochytrium dendrobatidis* (*Bd*)
BEAST. *See* Bayesian Evolutionary Analysis Sampling Trees (BEAST)
Beijing Genomics Institute (BGI), 430−432
Beta-2-microglobulin (b2m), 246
Bexsero, 214
BGI. *See* Beijing Genomics Institute (BGI)
BI. *See* Bayesian inference (BI)
Bias error. *See* Systematic error
BIC. *See* Bayesian Information Criterion (BIC)
"Big data", 405, 423
Biodiversity, 297−299
Biogeographical islands, 300−301
Biogeography, 367−368
Bioinformatics, 405, 413
Biological evolution, 167−168
Biological fluids, 234
Biological species, 5
Biomarkers linked to infection process by pathogen using SELDI-TOF-MS technology, 235−236
Bionominalism, 4−5
Bionominalists, 3
Biosphere, asexuality in, 103
Biotech companies, 423
BK virus (BKV), 155
BKV. *See* BK virus (BKV)
Blood feeding, 407−408
Borrelia burgdorferi (*B. burgdorferi*), 37
Bovine TB, 537
Breast cancer-related protein (BCRP/ABCG2), 436
Bs. *See Bacillus sphaericus* (Bs)
Bt. *See Bacillus thuringiensis* (Bt)
Bushmeat hunting, 587−588

C

CA-MRSA. *See* Community-associated methicillin-resistant *S. aureus* (CA-MRSA)
Cag pathogenecity island (*cag*PAI), 143, 149
California National Primate Research Center (CNPRC), 575
Canalization, 294−295
Candida albicans (*C. albicans*), 234−235, 471
Candida species, 73, 77
Candidate gene approach, 511−512, 516
Canine visceral leishmaniasis (CVL), 474
CAR. *See* Constitutive androstane receptor (CAR)
Carbamates (CXs), 315
Carboxylesterases (COEs), 319, 321−322
Cardiovascular disease (CVD), 435−436
Cassette chromosome recombinase gene complex (*ccr* gene complex), 554
CC398 strains, 564
CC97 strains, 564
CCC. *See* Chronic Chagas cardiomyopathy (CCC)
ccr gene complex. *See* Cassette chromosome recombinase gene complex (*ccr* gene complex)
CCs. *See* Clonal complexes (CCs)
CD. *See* Chagas disease (CD); Character displacement (CD); Cyclodiene (CD)
Cell immunobiology, 238−239
Cenancestor. *See* Last universal common ancestor (LUCA)
Central mystery of taxonomy, 8
Central nervous system (CNS), 437

Downloaded From Vet-eBooks.com

Centroid size (CS), 286, 287f
Cercocebys atys. *See* Sooty mangabeys (*Cercocebys atys*)
CETP. *See* Cholesteryl ester transfer protein (*CETP*)
CF. *See* Cystic fibrosis (CF)
cg MLST. *See* Core-genome MLST (cg MLST)
Chagas disease (CD), 341, 465–466, 509–510. *See also* Infectious disease genomics
  genetics, 351–361
    dramatic reduction of disease prevalence, 351
    vectors, 351–352
    zoonosis, 351
  integrated genetic epidemiology, 509
    cycle, 510
    host genetic susceptibility to, 510–516
    parasite genetic diversity, 517–521
    vector genetic diversity, 517
Chagasic cardiopathy, 509–510
Character displacement (CD), 293–294
Characters, 171–173
  coding, 176
Charadriiformes, 621, 630–632
Chemokine and chemokine receptor genes, 514
Chemotherapy
  evolution of parasite resistance
    in artemisinin combination therapy era, 499–500
    to CQ, 498–499
Chloroquine (CQ), 489
  evolution of parasite resistance to, 498–499
Cholesteryl ester transfer protein (*CETP*), 441
Chromosomal mutations, 272
Chronic Chagas cardiomyopathy (CCC), 510–511
Chytridiomycota, 71–72, 75–76
CL Brener
  haploid genome, 467
  strain, 466
"Clades", 633
"Class inclusion", 3–4
Class membership, 3
Classical HLA associations, 512
CLIA regulations. *See* Clinical Laboratory Improvement Amendments regulations (CLIA regulations)
CLIC. *See* Collection of Landmarks for Identification and Characterization (CLIC)
Clinical Laboratory Improvement Amendments regulations (CLIA regulations), 432
Clonal complexes (CCs), 151–154, 556
Clonal evolution
  asexual reproduction, 99–100
  asexuality in biosphere, 103
  clonal microevolution, 106–109
  clonal modes, 102–103
  evolution and paradox of sex, 105–106
  genetic consequences of asexuality, 103–105
  LD, 100–101
  origin of life, 101–102
  origin of propagation, 101–102
  recombination, 101–102
  sexual reproduction, 100
Clonal interference, 108
Clonal microevolution, 106–109
  neutral loci variability in clonal populations, 106–107
  selection and adaptation in clonal populations, 107–109
Clonal modes, 102–103
Clonal organisms, 390–391
ClonalframeML, 62
Clonality, 101–102
  threshold, 518–520
Cluster class. *See* Polythetic class
Cluster-based approach, 25
Clustering method, 85
CNPRC. *See* California National Primate Research Center (CNPRC)
CNS. *See* Central nervous system (CNS)
Coalescent genealogy samplers, 82
Coalescent theory, 82
COEs. *See* Carboxylesterases (COEs)
Coevolution
  comparisons between coevolving organisms across time, 124–127
  dynamics between hosts and pathogens, 115
  of host and pathogen, 115–116

Downloaded From Vet-eBooks.com

Coevolution (*Continued*)
implications
diversification and speciation, 130
maintenance of genetic diversity, 130−131
phenomena, 517
of *Plasmodium* species, 490
rate between populations of *P. fluorescens* and SBW25Φ2, 126−127, 126b
Cohesion, bacterial ecotypes, 41−42
Cohesive recombination model, 41
Collection of Landmarks for Identification and Characterization (CLIC), 296
Community-associated methicillin-resistant *S. aureus* (CA-MRSA), 553, 562
Comparative genomics, 405−408
of plant pathogens, 86−87
sequencing, 474−475
Compensatory mutations, 269
Complex life cycle, 142
Conceptual approaches to decipher host−parasite interactions, 236−239
Constitutive androstane receptor (CAR), 437, 442
Core-genome MLST (cg MLST), 387
Coronaviruses, 603−605
surveillance in wildlife animals, 614
CQ. *See* Chloroquine (CQ)
Cross-resistance, 328
Cross-species transmissions, 588−589
from other retroviruses from primates to humans, 590−591
Cryptococcus gattii (*C. gattii*), 73−74
evolution and emergence of pathogenic *C. gattii* genotypes, 74−77
global emergence of Amphibian Pathogen *Bd*, 75−77
Cryptococcus neoformans (*C. neoformans*), 73−74
Cryptococcus species, 77
Cryptosporidium (Apicomplexa), 168−169
CS. *See* Centroid size (CS)
Customary models in population geneticists, 203
CVD. *See* Cardiovascular disease (CVD)
CVL. *See* Canine visceral leishmaniasis (CVL)
CXs. *See* Carbamates (CXs)
Cyclical coevolution, 119
Cyclodiene (CD), 315
Cynomolgus macaque model, 542
CYP2A6, 442
CYP2B6 enzyme. *See* Cytochrome P450 2B6 enzyme (CYP2B6 enzyme)
CYP3A4 enzyme. *See* Cytochrome P450 3A4 enzyme (CYP3A4 enzyme)
CYPs. *See* Cytochrome P450 monooxygenases (P450)
Cystatin C (cysC), 246
Cystic fibrosis (CF), 38
Cysts, 410−411
Cytochrome P450 2B6 enzyme (CYP2B6 enzyme), 437
Cytochrome P450 3A4 enzyme (CYP3A4 enzyme), 437
Cytochrome P450 monooxygenases (P450), 319−320
Cytoeffectorium, 229
Cytogenetics, 341−344, 368
of triatominae, 354
Cytokine and cytokine receptor genes, 513−514
Cytokine IL-22, 123
Cytosensorium, 229

D
Darwin, Charles, 25
2-DE. *See* Two-dimensional electrophoresis (2-DE)
De novo genome assembly, 414
De novo pathogen discovery, 218
Deciphering of molecular strategies involved in parasite immune evasion, 233−234
Decision Theory (DT), 392
"Defensins", 236−237
Democratic Republic of Congo (DRC), 573
Demography, 197−198
transition, 208
Department of Energy (DOE), 211
Detoxification enzymes, 319
Developmental stability (DS), 294
Diachasmimorpha longicaudata (*D. longicaudata*), 292
Dicaryomycota, 71−72
Didelphis, 468−470
Differentiation, measures of, 84
Dihydroartemisinin, 442

Downloaded From Vet-eBooks.com

Dihydrofolate reductase (*folA*), 55
Dikaryotic fungi, 81
Dimorphic fungi, 73
Diploids, 81, 104–105
Discrete typing unit (DTU), 465–466, 518
Disease
  mating, and reproductive strategy, 206–207
  and standard of living in preindustrial societies, 202–203
Disease-resistance genes (R genes), 236–237
Disentangle host and parasite genome responses, 236–238
Dispersal, 81–83
  of antimicrobial agents, 258
Diversification, 130
DLST. *See* Double locus sequence typing (DLST)
DNA, 228
  barcoding, 9
  gyrase, 265–266
  repair genes, 266
  sequence, 9–10, 391, 393
DOE. *See* Department of Energy (DOE)
Dolutegravir (DTG), 441
Domestic birds
  HPAI viruses, 626–627
  LPAI viruses, 626–627
Double locus sequence typing (DLST), 556b–557b
"Down-regulatory" cytokine, 514
DRC. *See* Democratic Republic of Congo (DRC)
Drosophila, 236–237
Drug
  discovery, 214–215
  resistance, 216–217, 271–272, 499
    mutations, 271–272
  target, 215–216
  transporters, 445
DS. *See* Developmental stability (DS)
DT. *See* Decision Theory (DT)
DTG. *See* Dolutegravir (DTG)
DTU. *See* Discrete typing unit (DTU)
Duffy-negative chemokine receptor, 196

E

EAI. *See* East African Indian (EAI)
East African Indian (EAI), 150–151
Ebola virus disease, 408
eBURST approach. *See* Based Upon Related Sequences Types approach (eBURST approach)
EC. *See* Executive Committee (EC)
ECG. *See* Electrocardiography (ECG)
Ecology
  into bacterial systematics, 43–44
  ecological theory, 546
  ecology-blind approach, 29
Ecotypes, 26, 34, 43
  cohesive recombination model, 41
  dynamics of ecotype formation and periodic selection, 28f
  geotype plus Boeing model, 41
  putative ecotypes, 35–36
  recurrent niche invasion model, 40–41
EFA. *See* Elliptic Fourier algorithm (EFA)
Efavirenz (EFV), 435
Efficacy, 438, 444
EFV. *See* Efavirenz (EFV)
Electrocardiography (ECG), 511
Elliptic Fourier algorithm (EFA), 288
Elvitegravir (EVG), 441
Emerging plant diseases, 78–79
Entameoba invadens (*E. invadens*), 410–411
Entamoeba histolytica (*E. histolytica*), 410–411
Enterobacteriaceae, 261–262
Envelope (E) protein, 603–605
Environmental and community context, 122–123
Environmental variance of size *vs*. of shape, 300
Epidemiology, 79
Epigenome analysis, 429
Epitopes, 545
Eradication agenda, 501–502
ESAT-6, 540–541, 546–547
ESBL. *See* Extended-spectrum β-lactamases (ESBL)
Escherichia coli (*E. coli*), 214
  genomes, 56
Ester-resistance alleles, 322
Ethoxycoumarin, 320
Etravirine, 435, 439
Eukaryotes, 462–463

Downloaded From Vet-eBooks.com

Eurasian wigeons (*Anas penelope*), 624–625
EVG. *See* Elvitegravir (EVG)
Evolution(ary)
  of drug-resistant parasites
    in artemisinin combination therapy era, 499–500
    to CQ, 498–499
  genetics of avian influenza viruses
    ecological insights from evolutionary analysis, 629–633
    HPAI H5N1 Virus, 633
  networks, 184
  of *Plasmodium*
    chemotherapy and evolution of drug-resistant parasites, 498–500
    host transitions in speciation events within genus, 491
    importance of host specificity, 495–496
    P. falciparum, 491–492
    P. reichenowi, 491–492
    selection for immunological escaping variants, 501
    speciation between *P. ovale curtisi* and *P. ovale wallikeri*, 493–495
    zoonoses, 496–498
  responses. *See also* Infectious diseases
    agriculture, 198–199
    demography, 197–198
    disease, mating, and reproductive strategy, 206–207
    disease and standard of living in preindustrial societies, 202–203
    genetic responses to infectious disease in human, 196
    malaria, 199–202
    parasites, 197–198
    population limitation, 203–206, 205f
    prosperity and postindustrial era mortality decline, 207–208
  SIV, 580–581
  trees, 84–85, 181–182
ExaML, 61–62
Executive Committee (EC), 15
Experimental coevolution methods, 128
Extended-spectrum β-lactamases (ESBL), 259
Extensively drugresistant TB (XDR TB), 444–445
External barriers, 195

F

F-statistics, 81
FA. *See* Fluctuating asymmetry (FA)
FACS. *See* Fluorescence-activated cell sorting (FACS)
Falciparum malaria's unusual virulence, 200
FAS. *See* Fatty acid synthase (FAS)
FastQC tool, 413–414
FastTree, 61–62
Fatty acid synthase (FAS), 215
FGI. *See* Fungal Genome Initiative (FGI)
First fluorescence in situ hybridization (FISH), 368
Fitness, 269
  compensation, 271–273
  cost, 271–272
    in *M. tuberculosis*, 273–274
Fitness compensation, 271–273
Fluctuating asymmetry (FA), 295
Fluorescence-activated cell sorting (FACS), 411
folA. *See* Dihydrofolate reductase (*folA*)
Fourier's coefficients, 288
Frequent speciation models
  Nano-Niche model, 39–40
  Species-Less model, 39
  Speedy Speciation model, 38–39
Fungal Genome Initiative (FGI), 211–212
Fungal genomes, 87
Fungal infections, globally emerging, 74
Fungal pathogens. *See also* Pathogenic bacteria, population structure of
  fungal diseases, 72
  genomics of fungi, 86–88
  human and animal pathogenic fungi, 72–74
  modern molecular epidemiological tools, 79
  mycoses, 74–77
  plant diseases, 78–79
  plant pathogenic fungi, 77–78
  population genetics, 80–86
Fuzzy species, 51

G

G6PD deficiency. *See* Glucose-6-phosphatedehydrogenase deficiency (G6PD deficiency)
GABA receptor. *See* γ-Aminobutyric acid receptor (GABA receptor)

Downloaded From Vet-eBooks.com

GC. *See* Genomic content (GC)
GCP. *See* Good clinical practices (GCP)
GDA. *See* Gene duplication and amplification (GDA)
Gene duplication, 265, 539
Gene duplication and amplification (GDA), 272
Gene expression analysis, 409, 415–416
Gene flow, 81–83, 167–168, 186
  distribution of, 83
  and flow of migrants, 299–300
  rate and direction, 82–83
  across species boundaries, 57
Gene-D, 10–11
Gene-for-gene model (GFGM), 117b–118b
Gene-P, 10–11
"Genealogical nexus", 4–5
Generalized Procrustes analysis (GPA), 286–287
Generation time, 122
Genes coding for cytokines, 514
Genetic lineages. *See* Discrete typing unit (DTU)
Genetic(s)
  of adaptation, 269–271
  assimilation, 291
  capitalism, 259–260
  consequences of asexuality, 103–105
  diversity, 369
    maintenance, 130–131
    mitochondrial and nuclear diversities, 359t
    of triatominae, 354–360
  drift, 270, 290
  exchange in trypanosomatids, 459
    Leishmania, 474–476
    T. brucei, 459–465
    T. cruzi, 465–474
  of insect vectors
    A. gambiae complex, 361–367
    contributions of genetic studies of vectors, 341–343
    order Ixodida, 367–371
    significance and control of vector-borne disease, 341, 342f
    of triatominae and chagas disease, 351–361
    of tsetse flies and African trypanosomiasis, 343–351, 343f
  mapping, 369–370
  of metric change, 290–293
    genetic drift, 290
    heritability, 290–291
    hidden genetic variability, 291
    hybridism, 292–293
    shape as polygenic character, 290
  of order Ixodida, 367–371
  reassortment, 632–633
  responses to infectious disease in human, 196
  structure, 299
Genome mining, 215
Genome sequencing, 405–406
Genome-wide approach, 515–516
Genome-wide association studies (GWAS), 216–217, 515–516
Genomes OnLine Database (GOLD), 211–212
Genomic content (GC), 529
Genomic(s), 51–52, 369–370, 405, 421
  analysis tools for bacterial population structure, 61–63
  customers and needs, 428–430
    clinical biology and public health, 429
    other applications of genomics, 429–430
    research, 428–429
  of fungi
    comparative genomics of plant pathogens, 86–87
    comparing animal and plant pathogens, 87–88
  industry landscape, 430–433
  signals of selection due to host immunity, 490
  studies, 490
  technologies and instrument platforms
    nanopore-based sequencing, 426–427
    NGS based on SBS, 424–426
    Sanger sequencing, 423–424
    software, 427
Genotypes, 633
  to phenotype, 271–273
Geographical mosaic theory, 122
Geographical structure, 363
Geotype plus Boeing model, 41
GFGM. *See* Gene-for-gene model (GFGM)
GFP. *See* Green fluorescent protein (GFP)
Gilbert's syndrome, 440

Downloaded From Vet-eBooks.com

Global suppressors, 269
Glossina, 343–344
  G. brevipalpis, 351
Glossinidae family, 343–344
  Fusca group tsetse flies in Africa, 346f
  Palpalis group tsetse flies in Africa, 347f
  predicted distribution, 345f
GLP. *See* Good laboratory practices (GLP)
Glucose-6-phosphatedehydrogenase deficiency (G6PD deficiency), 199–200, 443
Glutathione S-transferases (GSTs), 319–320
Glycosylation, 228
GMP. *See* Good manufacturing practices (GMP)
GMSR. *See* Greater Mekong subregion (GMSR)
goeBURST, 391
GOLD. *See* Genomes OnLine Database (GOLD)
Good clinical practices (GCP), 432
Good laboratory practices (GLP), 432
Good manufacturing practices (GMP), 432
GPA. *See* Generalized Procrustes analysis (GPA)
GR. *See* Growth regulators (GR)
Gram-negative bacteria, 144
Greater Mekong subregion (GMSR), 216
Green fluorescent protein (GFP), 461
Growth regulators (GR), 316–317, 326–327
GSTs. *See* Glutathione S-transferases (GSTs)
Guangdong Province, 601
Gubbins method, 62
Gull influenza viruses, 630–632, 631f
GWAS. *See* Genome-wide association studies (GWAS)
gyrA gene, 272

H

H11N2 LPAI virus, 629–630
H37 strain, 540–541
H37Ra strain, 530
HA. *See* Hemagglutinin (HA)
HA-MRSA. *See* Hospital-associated methicillin-resistant *S. aureus* (HA-MRSA)
Haemophilus influenza (*H. influenza*), 211–212
Hardy–Weinberg equilibrium (HW equilibrium), 459–460
  genotypic proportions, 100
  and linkage equilibrium, 85
Harmonics, 288
HAT. *See* Human African trypanosomiasis (HAT)
HCV infection. *See* Hepatitis C virus infection (HCV infection)
HdeB proteins, 264–265
Helicobacter pylori (*H. pylori*), 144–150, 235
  population-genetic structure, 154
Hemagglutinin (HA), 621–623
  DNA maximum likelihood tree for HA, 630f
Hemiptera, 351–361
Hemophilus influenza (*H. influenza*), 421
Hepatitis C virus infection (HCV infection), 236
Hepatitis G virus (HGV), 155
Hepatotoxicity, 437, 445
Heritability, 290–291
Herpes simplex virus (HSV), 155
Heterogeneity in recombination, 54–55
Heuristic, 301
HGP. *See* Human Genome Project (HGP)
HGT. *See* Horizontal gene transfer (HGT)
HGV. *See* Hepatitis G virus (HGV)
Hidden genetic variability, 291
Hierarchical Likelihood Ratio Test (hLRT), 392
High-throughput sequencing, 413–414
Highly pathogenic avian influenza viruses (HPAI viruses), 623. *See also* Low pathogenic avian influenza viruses (LPAI viruses)
  in domestic birds, 626–627
  H5N1 virus, 627–629, 633
  H5NX virus, 627–629
Hill–Robertson effect, 108–109
HIV. *See* Human immunodeficiency virus (HIV)

Downloaded From Vet-eBooks.com

HIV therapy, pharmacogenetics off, 435–441. *See also* Antimalarial therapy, pharmacogenetics of
entry and integrase inhibitors, 441
NNRTIs, 437–439
NRTIs, 435–436
PIs, 439–441
HIV/AIDS, 573
HLA. *See* Human leukocyte antigen (HLA)
HLA B*5701, 436
hLRT. *See* Hierarchical Likelihood Ratio Test (hLRT)
HMP. *See* Human Microbiome Project (HMP)
Holistic approach, 236–238
Homo sapiens (*H. sapiens*), 494
hemoglobin structural variants, 490
Plasmodium phylogeny and description of species infecting, 487–489
Homology, 174, 285–286
Horizontal gene transfer (HGT), 87, 101, 257, 531
Hospital-associated methicillin-resistant *S. aureus* (HA-MRSA), 553, 563
Host
adaptation, 540, 543
coevolution, 115–116
antagonistic coevolution, 116–119
evolution of pathogen virulence, 119–120
implications of coevolution, 130–131
process of antagonistic coevolution, 121–124
testing for host–pathogen coevolution, 124–130
genetic susceptibility to CD, 510–516
genome responses, 236–238
genomic signals of selection due to host immunity, 490
proteome responses to parasite infection, 234–235
restriction factors, 589
specificity, 495–496
transitions, 493–495
in malaria parasites, 496
in speciation events within genus, 491
Host–parasite interactions, 227–228, 240–247
for parasites with simple or complex life cycle, 236–239
holistic approach, 236–238
pathogeno-proteomics, 238–239, 238f
Host–pathogen coevolution
testing for, 124
direct comparisons between coevolving organisms, 124–127
measuring population genetic change, 127–128
pathogen local adaptation, 129–130
pathogen-mediated rare host advantage, 128–129
of tubercle bacillus, 531–538
animal-adapted strains, 538
M. africanum, 535–536
M. bovis, 537
M. canettii, 536–537
M. tuberculosis, 535
MTBC members, 535
Host–pathogen interactions, 115, 227–229, 238–239, 239f
interest of proteomics, 228–229
microbiome influence on, 123
Host–vector–pathogen interactions, 238–239
HPAI viruses. *See* Highly pathogenic avian influenza viruses (HPAI viruses)
HPV. *See* Human papillomavirus (HPV)
HSRs. *See* Hypersensitivity reactions (HSRs)
HSV. *See* Herpes simplex virus (HSV)
HTLV. *See* Human T lymphotropic virus (HTLV)
Human(s)
and animal pathogenic fungi, 72–74
ascomycetes, 73
basidiomycetes, 73–74
globally emerging fungal infections, 74
cross-species transmissions in HIV-1 viruses in humans, 584
exposure to SIVs
cross-species transmissions, 588–591
host restriction factors, 589
human factors, 590
prevalences, 588–589

Downloaded From Vet-eBooks.com

Human(s) (*Continued*)
to SIV−infected nonhuman primates, 587−588
viral adaptation, 589−590
factors, 590
immune systems, 195
parasite
P. falciparum, 496−497
removal of species, 497
pathogens
harboring gene-transfer units, 264
origin of, 77
polyomavirus, 155−159
population proteomics, 241−247
SARS-CoV adaptation in, 610−612
SIV from chimpanzees and gorillas as ancestors, 581−584
SIV transmissions by exposure to infected primates, 585
Human African trypanosomiasis (HAT), 343
Human demography, microbes as tracers
candidates
bacteria, 144−155
viruses, 155−159
human evolutionary history, 141−142
pathogens as genetic tracers for host history, 142−143
Human Genome Project (HGP), 211
Human immunodeficiency virus (HIV), 212, 573. *See also* Acquired immune deficiency syndrome (AIDS)
and SIV, 575−585
from chimpanzees and gorillas, 581−584, 582f
cross-species transmissions in HIV-1 viruses, 584
diverging in human population, 585
genetic diversity and evolutionary history, 578f
infection, 235−236
transmissions to humans by exposure to infected primates, 585
Human leukocyte antigen (HLA), 435
Human Microbiome Project (HMP), 213
Human migrations, microbes as tracers of past
candidates
bacteria, 144−155
viruses, 155−159
human evolutionary history, 141−142
pathogens as genetic tracers for host history, 142−143
Human papillomavirus (HPV), 155
Human polyomavirus JC (JCV), 155
Human T lymphotropic virus (HTLV), 591
HTLV-1, 155
HTLV-2, 155
HW equilibrium. *See* Hardy−Weinberg equilibrium (HW equilibrium)
Hybrid vigor, 470
Hybridism, 292−293
Hybrids, 459−463
Hypersensitivity reactions (HSRs), 435

I

IBD model. *See* Isolation-by-distance model (IBD model)
ICTV. *See* International Committee on Taxonomy of Viruses (ICTV)
IDUs. *See* Intravenous drug users (IDUs)
ILD. *See* Inter-landmark distances (ILD)
ILS. *See* Incomplete lineage sorting (ILS)
Immune selection signals, 490
Immune system, 233
Incomplete lineage sorting (ILS), 393
Indinavir therapy, 439
Indoor residual spraying (IRS), 316
Industry landscape, 430−433
chronology of genomics companies creation, 431f
Infection genetics, 117b−118b
Infectious diseases, 53, 313, 405. *See also* Evolution(ary)—responses
big data analyzing, 405
comparative genomics, 405−408
de novo genome assembly, 414
genomics
clinical application, 218−219
drug discovery, 214−215
drug resistance, 216−217
drug target, 215−216
genome projects, 213t
history and development, 211
sequence diversity within species, 212
therapeutic response, 216−217
vaccine target, 214
vector control, 217
high-throughput sequencing, 413−414

Downloaded From Vet-eBooks.com

RNA-Seq, 415–416
single-cell technologies, 411–412
transcriptomics, 409–411
vectors, 234
WGS analysis, 414–415
Influenza A virus
classification, 623
and genetic variation of surface glycoproteins HA and NA, 622f
structure and genome organization, 621–623
taxonomy and host range, 621
wild waterfowl and shorebirds form reservoir, 622f
Influenza viruses
HPAI H5N1 and H5NX virus in wild birds, 627–629
LPAI virus
and highly pathogenic avian influenza viruses in domestic birds, 626–627
subtypes in wild birds, 624
transmission and epidemiology in wild birds, 624–626
Innate immune system, 195
Insect control, 217
Insecta, 172
Insecticide resistance, 314–318. *See also* Antibiotic resistance
alternative insecticides, 317–318
mechanisms of resistance, 318–328
synthetic insecticides, 315–317
Insecticide-treated nets (ITNs), 316
Insecticides, 314
Integrase inhibitors (InSTIs), 437, 441
Integrated genetic epidemiology, 509
of CD, 509–510
cycle, 510
host genetic susceptibility to, 510–516
parasite genetic diversity, 517–521
vector genetic diversity, 517
Integron, 263
Inter-landmark distances (ILD), 286
Interhost evolution and population structure, 59–61
Interleukin-1 receptor, 236–237
Internal transcribed spacer (ITS), 368–369
International Committee on Taxonomy of Viruses (ICTV), 1–2
definition of virus species, 11–14
International Union of Microbiologial Societies (IUMS), 1–2
Interspecies transmission, 613
Intrahost evolution, 58–59
Intravenous drug users (IDUs), 573
Intrinsic resistance, 261
Ion S5 system, 426
IRS. *See* Indoor residual spraying (IRS)
IS6110 element, 539–540
Isoenzyme diversity interpretation, 518
Isolation, 301
Isolation-by-distance model (IBD model), 141
Isoniazid toxicity, 445
Isothermal amplification approach, 411
Iterative Virus Assembler (IVA), 414
ITNs. *See* Insecticide-treated nets (ITNs)
ITS. *See* Internal transcribed spacer (ITS)
IUMS. *See* International Union of Microbiologial Societies (IUMS)
IVA. *See* Iterative Virus Assembler (IVA)
Ixodes, 407–408
I. ricinus, 367–368
I. scapularis, 407–408
Ixodida order, genetics of, 367–371
cytogenetics, 368
genetic diversity and population genetics, 369
genomics and genetic mapping, 369–370
perspectives for control and future directions, 371
phylogenetics and molecular diagnostics, 368–369
systematics, biogeography, and medical/veterinary significance, 367–368
ticks, 367

J

JCV. *See* Human polyomavirus JC (JCV)
JH. *See* Juvenile hormone (JH)
Juvenile hormone (JH), 316–317

K

K13-propeller polymorphism, 216–217
KD effect. *See* Knock-down effect (KD effect)
kDNA. *See* Kinetoplast DNA (kDNA)
kdr. *See* Knockdown resistance (kdr)

Downloaded From Vet-eBooks.com

Kinetoplast DNA (kDNA), 464
inheritance, 464
Kinetoplastidae family, 510
Klebsiella pneumonia (*K. pneumonia*), 257
Kluyvera ascorbata (*K. ascorbata*), 261–262
Knock-down effect (KD effect), 316
Knockdown resistance (kdr), 324, 365

L
LA. *See* Local adaptation (LA)
LA-MRSA. *See* Livestock-associated MRSA (LA-MRSA)
β-Lactamases, 267–269, 553
Landmark-based geometric morphometry, 285–288
semilandmarks for curves shape, 287–288
Landmark-based shape, 286–287, 296
Landmark-based size, 286
size variable, 286
Last universal common ancestor (LUCA), 101
LD. *See* Linkage disequilibrium (LD); Lyme disease (LD)
Legacy of habitat destruction for wild primates, 496–498
Legionella pneumophila (*L. pneumophila*), 415
Leishmania, 459
evidence of genetic exchange in natural populations, 474–475
genetic crosses, 475–476
implications for epidemiology, 476
L. amastigotes, 233–234
L. infantum, 474
Lentivirus, 575
Leukocidin DE (*lukDE*), 560
Likelihood ratio tests (LRT), 394
Lineages, 531–536, 542–543, 546–547
Linkage disequilibrium (LD), 474–475, 511–512
Lipooligosaccharide (LOS), 536–537
Livestock-associated MRSA (LA-MRSA), 564
Living organisms, 227
Local adaptation (LA), 117b–118b
Logic of phylogeny reconstruction, 169–171
LOH. *See* Loss of heterozygosity (LOH)
Long-Read SBS, 426
Lopinavir/ritonavir (LPV/r), 435
LOS. *See* Lipooligosaccharide (LOS)
Loss of heterozygosity (LOH), 76
Low pathogenic avian influenza viruses (LPAI viruses), 623. *See also* Highly pathogenic avian influenza viruses (HPAI viruses)
and highly pathogenic avian influenza viruses in domestic birds, 626–627
subtypes in wild birds, 624
transmission and epidemiology in wild birds, 624–626
LPAI viruses. *See* Low pathogenic avian influenza viruses (LPAI viruses)
LPV/r. *See* Lopinavir/ritonavir (LPV/r)
LRT. *See* Likelihood ratio tests (LRT)
LTA. *See* Lymphotoxin-alpha (LTA)
LUCA. *See* Last universal common ancestor (LUCA)
lukDE. *See* Leukocidin DE (*lukDE*)
Lumefantrin, 442–443
Lyme disease (LD), 53, 367
Lymphotoxin-alpha (LTA), 512–513

M
Macaca arctoides. *See* Stump-tailed macaques (*Macaca arctoides*)
Major facilitator superfamily (MFS), 86–87
Major histocompatibility complex (MHC), 198–199
associations, 512–513
Malaria, 207, 406–407, 442, 444
lessons from, 199–202
malaria transmission and control, *A. gambiae* population genetics to, 366
malaria vaccines, selection for immunological escaping variants by, 501
MALDI. *See* Matrix–assisted laser desorption ionization (MALDI)
Malthusian population, 204
MAM. *See* Matching alleles model (MAM)
Mannose-binding lectin (MBL), 515
Maraviroc (MVC), 435, 441

Downloaded From Vet-eBooks.com

Markov chain Monte Carlo approach (MCMC approach), 157–158, 393
Mass spectrometric immunoassay (MSIA), 242, 243f
Mass spectrometry (MS), 229
Matching alleles model (MAM), 117b–118b
Mating system, 72, 80–81
Matrix–assisted laser desorption ionization (MALDI), 235
Maxicircle kDNA genotypes, 475
Maximum likelihood (ML), 61–62
  inference, 392
  methods, 61–62
MBL. *See* Mannose-binding lectin (MBL)
MCC. *See* Merkel cell carcinoma (MCC)
MCL. *See* Mucocutaneous leishmaniasis (MCL)
MCMC approach. *See* Markov chain Monte Carlo approach (MCMC approach)
MCV. *See* Merkel cell polyomavirus (MCV)
MD. *See* Metric disparity (MD)
MDR. *See* Multidrug resistance (MDR)
MDS. *See* Multidimensional scaling (MDS)
Measles virus, 2–4, 15–16
Measurement error, 289
mec gene complex (*mec* gene complex), 554
mecA gene homologue (*mecC*), 555–556
Median network, construction and interpretation, 187f
Medical entomology applications
  biodiversity, 297–299
  characterization tool, 297
  correct species attribution scores based, 298t
  morphometric identification, 297t
  population structure, 299–301
  reinfestation studies, 299
  species identification and detection, 296
Medical/veterinary significance, 367–368
Mefloquine, 443–444
Meiosis, 462–464
Meloidogyne, 236–237
Membrane (M) protein, 603–605
MenB. *See* Serogroup B (MenB)
Mendelian inheritance, 462–464
Mendelian rules of inheritance, 459
Merkel cell carcinoma (MCC), 218–219
Merkel cell polyomavirus (MCV), 218–219
MERS. *See* Middle East Respiratory Syndrome (MERS)
Meselson effect, 104–105, 104f
Mesoamerica, 76
Metabolic resistance, 319–322
  carboxylesterases, 321–322
  GSTs, 319–320
  P450, 320–321
Metabolic toxicity, 435
Metagenomic study, 37
Metarhizium anisopliae (*M. anisopliae*), 317–318
Methicillin-resistant *Staphylococcus aureus* (MRSA), 52–53, 259, 395–396, 555–556
  clones, 554–555
  evolution, 556–560
  genetic diversity, 559
  MGEs, 559–560
  molecular epidemiology, 561–564
Metric disparity (MD), 298
MFS. *See* Major facilitator superfamily (MFS)
MGEs. *See* Mobile genetic elements (MGEs)
MGP. *See* Microbial genome program (MGP)
MHC. *See* Major histocompatibility complex (MHC)
MIC. *See* Minimal inhibitory concentration (MIC)
Microarrays, 427
  assay, 218
  technologies, 421–422
Microbes as tracers of past human demography and migrations
  candidates
    bacteria, 144–155
    viruses, 155–159
  human evolutionary history, 141–142
  pathogens as genetic tracers for host history, 142–143
Microbial genome program (MGP), 211–212
Microbiology, 25–26
Microbiome influence on host–pathogen interactions, 123
Microcell, 409–410

Downloaded From Vet-eBooks.com

Microsatellite loci, genetic variability, 344, 348t, 364f
  mitochondrial diversities and genetic differentiation, 349t
Middle East Respiratory Syndrome (MERS), 614
Migration, 81–83, 121–122, 624–626, 629–630
Mikrocytos mackini (*M. mackini*), 409–410
Minimal inhibitory concentration (MIC), 257–258
MIRUs. *See* Mycobacterial interspersed repetitive units (MIRUs)
Mismatch repair system (MMR), 266
Mites, 367–369
ML. *See* Maximum likelihood (ML)
MLEE. *See* Multilocus enzyme electrophoresis (MLEE)
MLGs. *See* Multilocus genotypes (MLGs)
MLMT. *See* Multilocus microsatellite typing (MLMT)
MLST. *See* Multilocus sequence typing (MLST)
MLVA. *See* Multilocus Variable number of tandem repeats Analysis (MLVA)
MMR. *See* Mismatch repair system (MMR)
Mobile genetic elements (MGEs), 259, 274–275, 553–554, 559–561
Model-based Bayesian clustering algorithms, 85
Molecular biologists, 228
Molecular characters, 172
Molecular clock analysis, 60
Molecular diagnostics, 368–369
Molecular epidemiology, 395
  MLST application, 395–396
  of MRSA, 561–564
Molecular evolution of SARS-CoV, 610–613
  rapid adaptation in humans, 610–612
  receptor usage and evolutionary selection, 612–613
  viral genetic diversity generation by recombination, 612
Molecular evolution rate, 143
Mononegavirales, 1
Monothetic classes, 1–2, 7f
Morbillivirus, 3–4
Morphometrics
  modern and traditional, 285
  revolution, 285
Mosquito vectors, 407–408
Most recent common ancestor (MRCA), 143, 534, 610–612
MRCA. *See* Most recent common ancestor (MRCA)
MRP4/ABCC4. *See* Multidrug resistance protein 4 (MRP4/ABCC4)
MRP5/ABCC5. *See* Multidrug-resistant protein 5 (MRP5/ABCC5)
MRSA. *See* Methicillin-resistant *Staphylococcus aureus* (MRSA)
MS. *See* Mass spectrometry (MS)
MSA. *See* Multiple sequence alignment (MSA)
MSIA. *See* Mass spectrometric immunoassay (MSIA)
MTBC. *See Mycobacterium tuberculosis* complex (MTBC)
mtDNA, genetic variability, 344
Mucocutaneous leishmaniasis (MCL), 474
Multi-Locus Strain Typing (MLST). *See* Multilocus sequence typing (MLST)
Multidimensional scaling (MDS), 151–154
Multidrug resistance (MDR), 54, 260–261
  TB, 444–445
Multidrug resistance protein 4 (MRP4/ABCC4), 436
Multidrug-resistant protein 5 (MRP5/ABCC5), 436
Multilocus enzyme electrophoresis (MLEE), 383, 465–466, 556b–557b
Multilocus genotypes (MLGs), 103–104
Multilocus microsatellite typing (MLMT), 79, 471
Multilocus sequence typing (MLST), 51, 79, 383, 414, 426, 473, 556b–557b, 559–560, 563
  databases, 388
  of pathogens
    advantages and disadvantages, 389–390
    analytical approaches, 390–394
    applications, 395–397
    molecular design and development of MLST, 384–387

Downloaded From Vet-eBooks.com

Multilocus Variable number of tandem repeats Analysis (MLVA), 387
Multiple mutations model, 108
Multiple sequence alignment (MSA), 392
Multivariate methods, 85
Mustached monkeys (SIVmus), 580–581
Mutation, 54, 58–59, 121–122, 261
  rate, 125–126
MVC. *See* Maraviroc (MVC)
Mycobacteria, 529
  disease and mycobacterial genetics, 529–531
Mycobacterial interspersed repetitive units (MIRUs), 150
  MIRU-VNTR markers, 150–151
Mycobacterium species, 529, 531
  catalase gene, 261
  M. africanum, 535–536
  M. bovis, 537
  M. canettii, 536–537
  M. caprae, 538
  M. pinnipedii, 538
  M. smegmatis, 267
Mycobacterium tuberculosis (*M. tuberculosis*), 150–155, 257, 529, 535
  adaptive cues of, 544–547
  epidemiological and experimental fitness cost, 273–274
  evolution in laboratory environment, 540–541
  evolution of, 538–540
  short-term evolution of, 542–544
  in vitro attenuation of bacteria, 540–541
Mycobacterium tuberculosis complex (MTBC), 150, 529
  deletion-based phylogeny, 532f–533f
  genome-based phylogeny of mirrors, 152f
Mycoplasma genitalium (*M. genitalium*), 211–212
Mycoses
  evolution and emergence of pathogenic *C. gattii* genotypes, 74–77
  origin of human pathogens, 77

N

NA. *See* Neuraminidase (NA)
Nalidixic acid, 272
Nano-Niche
  ecotypes, 39
  model, 39–40
Nanopore-based sequencing, 426–427
  microarrays, 427
NAT2. *See* N-Acetyltransferase 2 (NAT2)
National Human Genome Research Institute (NHGRI), 211, 422–423
National Institutes of Health (NIH), 211
National Research Council (NRC), 211
Natural reservoirs, 629–633
  of SARS-CoV, 606–610
Natural selection, 257, 270–271
NC-motif. *See* Nucleotide combination motif (NC-motif)
Neighbor joining (NJ), 392
Neisseria spp., 52–53
  N. meningitides, 395–396
Nelfinavir, 440
NEP. *See* Nucleic export protein (NEP)
NERPRC. *See* New England Regional Primate Research Center (NERPRC)
Neuraminidase (NA), 621–623
Nevirapine (NVP), 435, 437
New England Regional Primate Research Center (NERPRC), 575
New World primates (NWPs), 590–591
Next-generation sequencers, 413
Next-generation sequencing (NGS), 405, 421–422. *See also* Infectious diseases
  based on SBS, 424–426
    short-read sequencing, 424–426
    single-molecule sequencing, 426
  technologies, 219, 367
Next-generation SOLiD sequencing, 410–411
NGS. *See* Next-generation sequencing (NGS)
NHGRI. *See* National Human Genome Research Institute (NHGRI)
NHP. *See* Nonhuman primates (NHP)
NIH. *See* National Institutes of Health (NIH)
NJ. *See* Neighbor joining (NJ)
NNRTIs. *See* Nonnucleoside reverse transcriptase inhibitors (NNRTIs)
Non-Latinized binomial names for virus species, 14–16
Nonclonal organisms, 390–391

Downloaded From Vet-eBooks.com

Nondimensional structure, 363–365
Nonhuman primates (NHP), 575
Nonmodel-based approaches, 61
Nonnucleoside reverse transcriptase inhibitors (NNRTIs), 437–439
Nonstructural protein 3 (Nsp 3), 607–610
Northwestern population group (NW population group), 363
NP. *See* Nucleocapsid protein (NP)
NRC. *See* National Research Council (NRC)
NRTIs. *See* Nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs)
NS1, 621–623
NS2, 621–623
Nsp 3. *See* Nonstructural protein 3 (Nsp 3)
Nuclear receptors, 445
Nucleic export protein (NEP), 621–623
Nucleocapsid protein (NP), 603–605, 621–623
Nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs), 435–436
Nucleotide
  bias, 178
  nucleotide-based methods, 390–394
    phylogenetic relatedness, 392–393
    population dynamics, 393–394
Nucleotide combination motif (NC-motif), 11–12
Nullarbor, 415
NVP. *See* Nevirapine (NVP)
NW population group. *See* Northwestern population group (NW population group)
NWPs. *See* New World primates (NWPs)

O

Ochrobactrum anthropi (*O. anthropi*), 395
OCs. *See* Organochlorines (OCs)
Office of Technology Assessment (OTA), 211
"Omic" approaches, 531
OMVs. *See* Outer-membrane vesicles (OMVs)
One-by-one mutation model, 108
"Ontogenetic" allometry, 289
Oomycetes, 71–72, 78
Open reading frames (ORFs), 214, 603–605
OPs. *See* Organophosphates (OPs)
Oral Polio Vaccine (OPV), 584
ORF8 coding region, 612
ORFs. *See* Open reading frames (ORFs)
Organochlorines (OCs), 315
Organophosphates (OPs), 315
Origin of life, 101–102
Origin of propagation, 101–102
ORM1. *See* Orosomucoid 1 (ORM1)
Orosomucoid 1 (ORM1), 439–440
Orthomyxoviridae, 621
OTA. *See* Office of Technology Assessment (OTA)
Outer-membrane vesicles (OMVs), 214
"Outlier detection" methods, 63
Outline-based
  shape, 288
  size, 288
Oyster pathogens, 409–410

P

P450. *See* Cytochrome P450 monooxygenases (P450)
PA. *See* Polymerase (PA)
PA-X protein, 621–623
Pacific Northwest, evolution and emergence of pathogenic *C. gattii* genotypes, 74–77
Pairwise Sequence Comparison (PASC), 13–14
Pan-genome of species and populations, 55–57
Panmixia, 100
Panton-Valentine leukocidin toxin (PVL toxin), 562–563
Paradox, 234
Parasite(s), 197–198
  genetic diversity, 517–521
  genome responses, disentangle, 236–238
  host–parasite interactions for with simple or complex life cycle, 236–239
  immune evasion, 233–234
    deciphering of molecular strategies in, 233–234
  infection, host proteome responses to, 234–235
  parasite-clearance rate reduction, 216
Parasito-proteomics, 236–237
Partial warps, 286–287
PASC. *See* Pairwise Sequence Comparison (PASC)

Downloaded From Vet-eBooks.com

Pathogen(s), 383–384, 389–390, 407–408. *See also* Phylogenetic analysis
coevolution, 115–116
antagonistic coevolution, 116–119
evolution of pathogen virulence, 119–120
implications of coevolution, 130–131
process of antagonistic coevolution, 121–124
testing for host–pathogen coevolution, 124–130
dynamics, 197–198
as genetic tracers for host history, 142–143
local adaptation, 129–130
pathogen-mediated rare host advantage, 128–129
using SELDI-TOF-MS technology, 235–236
virulence evolution, 119–120
Pathogenesis, 540
Pathogenic bacteria, population structure of. *See also* Fungal pathogens
bacterial species, 51
genomic analysis tools, 61–63
intrahost evolution, 58–59
and population structure, 59–61
recombination in bacterial populations, 52–57
Pathogenic *Cryptococcus* species complex, 73–74
Pathogenic diseases, 227–228
Pathogenicity of SIVs in natural hosts, 579–580
Pathogeno-proteomics, 238–239, 238f
PB1. *See* Basic polymerase 1 (PB1)
PB1-F2, 621–623
PCA. *See* Principal Component Analysis (PCA)
PCE. *See* Predominant clonal evolution (PCE)
PCR. *See* Polymerase chain reaction (PCR)
PCR-based approach, 411
PDIM. *See* Phtiocerol dimycocerosate (PDIM)
Penicillin-resistant staphylococci, 553
Penicillinase, 553
PEP. *See* Phosphoenolpyruvate (PEP)
Periodic selection, 26–28
model, 108
PFGE. *See* Pulsed-field gel electrophoresis (PFGE)
Pharmacogenetics of infectious disease therapy
antimalarial therapy, 442–444
antituberculous therapy, 444–445
HIV therapy, 435–441
Phenotype, 10–11, 285, 510–511
genotype to, 271–273
phenotypic analysis, 475
regulation, 294–295
canalization, 294–295
developmental stability, 295
Phenotypic plasticity (PP), 293
Phlebotomus duboscqi (*P. duboscqi*), 475
PhoP-PhoR virulence system, 540–541
Phosphoenolpyruvate (PEP), 215
Phospholipase D (PLD), 410–411
Phosphorylation, 228
Phtiocerol dimycocerosate (PDIM), 540–541
Phylogenetic(s), 368–369
coalescent models, 396
MLST application, 396–397
tree, 167f
of betacoronavirus group b, 603–605, 605f
Phylogenetic analysis, 75, 415. *See also* Pathogen(s)
characteristics of contemporary organisms, 167
characters and samples, 171–173
choosing method, 179–181
logic of phylogeny reconstruction, 169–171
phylogenetic networks, 184–188
phylogenetic tree, 167f
practice of phylogeny reconstruction, 174–179, 175f
trees, 181–184, 182f
types of biological evolution, 167–168
uses of phylogenies, 168–169
Phylogeny, 157
and description of species infecting *Homo Sapiens*, 487–489
genome-based, 152f
M. tuberculosis, 150–151
MTBC, 151
reconstruction, logic of, 169–171

Downloaded From Vet-eBooks.com

Phylogeny (*Continued*)
reconstruction practice, 174−179, 175f
representation, 181−184
SIV, 580−581
uses, 168−169
PhyML, 61−62
Pilus, 100
PIs. *See* Protease inhibitors (PIs)
Plant diseases, 78−79
Plant pathogens. *See also* Pathogen(s)
comparative genomics of, 86−87
plant pathogenic fungi, 77−78
Plasmodium, 313−314, 487. *See also* Infectious disease genomics
design of public health interventions, 489
evolution
21st century in three courses, 496−501
and eradication agenda, 501−502
last 10 million years, 491−496
genomic signals of selection due to host immunity, 490
genomic studies, 490
P. falciparum, 178, 200, 442, 487, 491−492
P. falciparum malaria, 442
P. knowlesi, 487
zoonosis emergence, 497−498
P. malariae, 487−489
P. ovale curtisi, 493−495
P. ovale wallikeri, 493−495
P. reichenowi, 491−492
P. vivax, 200
phylogeny and description of species infecting *Homo sapiens*, 487−489
population genetics, 489
PLD. *See* Phospholipase D (PLD)
PMTCT. *See* Prevention of mother to child transmission (PMTCT)
Pneumococcal surface protein A (*pspA*), 55
Pneumococcal surface protein C (*pspC*), 55
Polygenic character, shape as, 290
Polymerase (PA), 621−623
Polymerase chain reaction (PCR), 215
Polythetic class, 1−2, 7−9, 7f
Population
biogeographical islands, 300−301
environmental variance of size *vs*. of shape, 300
gene flow and flow of migrants, 299−300
limitation, 203−206, 205f
need for heuristic, 301
of pathogenic bacteria
bacterial species, 51
genomic analysis tools, 61−63
interhost evolution and population structure, 59−61
intrahost evolution, 58−59
recombination in bacterial populations, 52−57
proteomics, 240−247, 241f
human population proteomics, 241−247
for living organisms, 240−241
structure, 299−301
subdivision, 83−85
evolutionary trees, 84−85
measures of differentiation, 84
model-based Bayesian clustering algorithms, 85
multivariate methods, 85
of triatominae, 360
of tsetse flies, 350
"Population assignment" approach, 61
Population genetic(s), 369, 370f, 518. *See also* Proteomics
and design of public health interventions, 489
measuring change, 127−128
of pathogenic fungi, 80−86
dispersal, 81−83
empirical population genetics, 85−86
gene flow, 81−83
migration, 81−83
population subdivision, 83−85
reproductive system, 80−81
structure in *A. gambiae*, 363−365
geographical structure, 363
nondimensional structure, 363−365
temporal structure, 363
studies, 71, 76
Post-genomic technology, 229
Postindustrial era mortality decline, 207−208
PP. *See* Phenotypic plasticity (PP)
Pragmatic approach, 411
Predominant clonal evolution (PCE), 517, 521
Pregenomic studies, 530−531
Pregnane X receptor (PXR), 440

Downloaded From Vet-eBooks.com

Preindustrial societies
disease and standard of living in, 202−203
Prevention of mother to child transmission (PMTCT), 437
Primaquine, 443
Principal Component Analysis (PCA), 61
Procrustes
paradigm, 287−288
residuals, 286−287
Profiles database, 388
Proguanil, 444
Prokaryotes, 102
Prosperity era mortality decline, 207−208
Protease inhibitors (PIs), 437, 439−441
Protein(s)
adaptive evolution of proteins encoding, 267−268
protein-coding DNA sequences, 394
Proteomics, 227−228, 231t−232t. *See also* Population genetic(s)
conceptual approaches to decipher host−parasite interactions, 236−239
to host−pathogen interactions, 228−229
population proteomics, 240−247, 241f
proteomic-based amniotic fluid analysis, 234−235
retrospective analysis, 229−236
Protists, 464
Pseudo-landmark-based shape, 288
outline-based shape, 288
outline-based size, 288
Pseudomonas fluorescens (*P. fluorescens*), 126b
Pseudomonas syringae (*P. syringae*), 116
pspA. *See* Pneumococcal surface protein A (*pspA*)
pspC. *See* Pneumococcal surface protein C (*pspC*)
Public health, MLST application, 395−396
PubMLST site, 388
Pulsed-field gel electrophoresis (PFGE), 556b−557b
Putative ecotypes
ecological diversity within, 36−38
from sequence data, 35−36
PVL toxin. *See* Panton-Valentine leukocidin toxin (PVL toxin)
PXR. *See* Pregnane X receptor (PXR)
Pyrethroids (PYRs), 315

Q
Quality of human WGS, 421
Quantitative proteomics strategies, 229, 230f
Quantitative trait loci (QTL), 290, 319−320, 370
Quinine, 444
Quinoneimines (QNMs), 443

R
R genes. *See* Disease-resistance genes (R genes)
r/m. *See* Relative to mutation (r/m)
RADseq technique. *See* Restriction site−associated DNA sequencing technique (RADseq technique)
RAL. *See* Raltegravir (RAL)
Raltegravir (RAL), 435, 441
Random error. *See* Stochastic error
Randomizations, 157−158
Randomly amplified polymorphic DNA (RAPD), 172, 471
RAST tool, 414−415
RAxML, 61−62
RBD. *See* Receptor-binding domain (RBD)
RBP. *See* Retinol-binding protein (RBP)
rDNA. *See* Ribosomal DNA (rDNA)
RDP4software packages, 393−394
REA. *See* Restriction endonuclease analysis (REA)
Reactive oxygen species (ROS), 266
Reads, 411−415
Real-Time SBS, 426
Receptor-binding domain (RBD), 612−613
Recombination, 26−28, 32−34, 100−102, 121−122
in bacterial populations
emergence and persistence of sequence clusters, 52−54
gene flow across species boundaries, 57
heterogeneity in recombination, 54−55
pan-genome of species and populations, 55−57
cohesive recombination model, 41
network, 185f
rate, 143
Recurrent niche invasion model, 40−41
Red fluorescent protein (RFP), 461

Downloaded From Vet-eBooks.com

"Red Queen"
dynamics, 117b–118b, 119, 128–130
hypothesis, 101, 105–106
Reduviidae, 351–361
Reference images, need for bank of, 296
Reinfestation studies, 299
Relative genetic contribution method (rGC method), 435
Relative to mutation (r/m), 52–53
Relative warps, 286–287
Reproductive isolation (RI), 365
Reproductive system, 80–81
Reservoir hosts, 625
Resistance, 314–315
generalities, 327–328
growth regulators, 326–327
mechanisms, 318–328
metabolic resistance, 319–322
mutation, 269–270
target-site modification, 322–326
toxin receptors, 327
Resistance ratio (RR), 314–315
Resorufin, 320
Restriction endonuclease analysis (REA), 396–397
Restriction fragment length polymorphism (RFLP), 172
Restriction site–associated DNA sequencing technique (RADseq technique), 369
Retinol-binding protein (RBP), 246
Retrospective analysis of previous proteomics studies, 229–236
biomarkers linked to infection process, 235–236
deciphering of molecular strategies, 233–234
host proteome responses to parasite infection, 234–235
Reversible termination SBS, 424–426
RFLP. *See* Restriction fragment length polymorphism (RFLP)
RFP. *See* Red fluorescent protein (RFP)
rGC method. *See* Relative genetic contribution method (rGC method)
Rhizobium leguminosarum (*R. leguminosarum*), 40
Rhodnius pallescens (*R. pallescens*), 293
Rhodnius prolixus (*R. prolixus*), 354
RI. *See* Reproductive isolation (RI)
Ribosomal DNA (rDNA), 368–369
Ribosomal Multilocus Sequence Typing method (rMLST method), 387
Rifampicin, 272, 445
Rilpivirine, 439
Ring-stage survival assay, 216
rMLST method. *See* Ribosomal Multilocus Sequence Typing method (rMLST method)
RMSF. *See* Rocky Mountain spotted fever (RMSF)
RNA
genome, 1
polymerase, 543–544
RNA-Seq, 415–416
sequence, 9–10
transcripts, 238–239
Rocky Mountain spotted fever (RMSF), 367
ROS. *See* Reactive oxygen species (ROS)
rpoB gene, 272
rpsL gene, 272
RR. *See* Resistance ratio (RR)
Ruddy turnstones (*Arenaria interpres*), 625
"Russian doll" pattern, 520f, 521

S

S-OIV. *See* Swine-origin influenza A virus (S-OIV)
S. zeamais primary endosymbiont (SZPE), 120
SAI. *See* Species-as-individuals (SAI)
Salmonella enteric (*S. enteric*), 272
SAMHD1 antiretroviral protein, 589
Samples, 171–173
Sand fly vectors, 468–470, 475
Sarcocystis, 169
SARS. *See* Severe acute respiratory syndrome (SARS)
SARS coronavirus (SARS-CoV), 603–605, 604f
animal link, 605–606
coronavirus surveillance in wildlife animals, 614
gene product comparison with bat SL-CoV, 608t–609t
molecular evolution, 610–613
natural reservoirs, 606–610

Downloaded From Vet-eBooks.com

SARS-like coronaviruses (SL-CoVs), 607
gene product comparison with SARS-CoV, 608t–609t
HKU3–8, 612
SBS. *See* Sequencing-by-synthesis (SBS)
SCC*mec* elements. *See* Staphylococcal chromosome cassette *mec* elements (SCC*mec* elements)
SCCs. *See* Staphylococcal chromosome cassettes (SCCs)
SCID. *See* Severe combined immunodeficiency (SCID)
Semiconductor sequencing, 426
Semilandmarks for curves shape, 287–288
Sequence alignment, 176
Sequence clusters
cohesive recombination model, 41
emergence and persistence of, 52–54
geotype plus Boeing model, 41
recurrent niche invasion model, 40–41
Sequence type (ST), 53, 383, 387, 556, 556b–557b, 558f, 561
Sequence Type Analysis and Recombinational Tests (START2), 391
Sequence Typing Analysis and Retrieval System (STARS), 391
Sequence-based algorithms, 36
Sequence-specific oligonucleotide (SSO), 512
Sequence-specific primer (SSP), 512
Sequencing, 429
Sequencing-by-synthesis (SBS), 421–422
NGS based on, 424–426
Serine, 613
Serogroup B (MenB), 214
Severe acute respiratory syndrome (SARS), 218, 601
chronological events, 602t
outbreaks, 603
Severe combined immunodeficiency (SCID), 470
Sex, evolution and paradox of, 105–106
Sexual recombination (SR), 100
Sexual reproduction, 100
SFV. *See* Simian foamy virus (SFV)
Shape
environmental variance, 300
as polygenic character, 290
Shewanella algae (*S. algae*), 261–262
Short-read sequencing, 424–426
Sign epistasis, 272
SILAC. *See* Stable isotope labeling by amino acids in cell culture (SILAC)
Simian foamy virus (SFV), 590–591
Simian immunodeficiency virus (SIV), 575
in African non-human primates, 575–579
discovery, 575
evolution and phylogeny, 580–581
exposure of humans to SIVs, 587–591
HIV-1 from, 581–585
HIV-2 from, 585–587, 587f
infections in Old World monkeys and apes in Africa, 576t–577t
pathogenicity in natural hosts, 579–580
Single nucleotide polymorphism (SNP), 58, 141, 369, 421, 558–559, 561
Single polypeptide precursor (HA0), 623
Single-cell technologies, 411–412
Single-molecule real-time sequencing technology (SMRT technology), 423
Single-molecule sequencing, 426
Single-nucleotide polymorphism (SNP), 511–512, 531–534
Single-nucleotide variations (SNVs), 610
Sitophilis zeamais (*S. zeamais*), 120
SIV. *See* Simian immunodeficiency virus (SIV)
SIVgsn. *See* Spot-nosed monkeys (SIVgsn)
SIVmac, 575
SIVmus. *See* Mustached monkeys (SIVmus)
Size, environmental variance, 300
Size variable, 286
SL-CoVs. *See* SARS-like coronaviruses (SL-CoVs)
Sliding landmarks, 287–288
SmaI enzyme, 556b–557b
SMRT technology. *See* Single-molecule real-time sequencing technology (SMRT technology)
SNP. *See* Single nucleotide polymorphism (SNP); Single-nucleotide polymorphism (SNP)
SNVs. *See* Single-nucleotide variations (SNVs)
Sooty mangabeys (*Cercocebys atys*), 575
SP. *See* Sulfadoxine–pyrimethamine (SP)
Spa-typing, 556b–557b

Downloaded From Vet-eBooks.com

Speciation, 25–26, 29, 71–72, 130
events within genus, host transitions in, 491
between *P. ovale curtisi* and *P. ovale wallikeri*, 493–495
Species
diagnosis, MLST application, 396–397
gene flow across boundaries, 57
identification and detection, 296
Species richness (SR), 298
Species-as-individuals (SAI), 4–5
Species-Less model, 39
of bacterial diversification, 42f
Speedy Speciation model, 38–39
Spike (S) protein, 603–605
Spot-nosed monkeys (SIVgsn), 580–581
SR. *See* Sexual recombination (SR); Species richness (SR)
SSE. *See* Super spread event (SSE)
SSO. *See* Sequence-specific oligonucleotide (SSO)
SSP. *See* Sequence-specific primer (SSP)
ST. *See* Sequence type (ST)
ST239-MRSA, 53
Stable Ecotype model, 37–38
of bacterial speciation, 33–35
Stable isotope labeling by amino acids in cell culture (SILAC), 233–234
Stable mutators, 266
Staphylococcal chromosome cassette *mec* elements (SCC*mec* elements), 553–556, 555t, 560
Staphylococcal chromosome cassettes (SCCs), 559–560
Staphylococcus aureus (*S. aureus*), 257, 395–396, 553
evolution, 556–560
SCC*mec* elements, 553–556, 555t
typing methods, 556b–557b
STARS. *See* Sequence Typing Analysis and Retrieval System (STARS)
START2. *See* Sequence Type Analysis and Recombinational Tests (START2)
Static allometry, 289
Stenotrophomonas maltophilia (*S. maltophilia*), 262–263
Stochastic error, 180
Stochastic processes, 270–271
Streptococcus agalactiae (*S. agalactiae*), 216
Streptococcus pneumoniae (*S. pneumoniae*), 429
Streptomyces platensis (*S. platensis*), 216
Streptomycin, 272
Stress, 295
STRUCTURE approach, 61
Stump-tailed macaques (*Macaca arctoides*), 575
Sub-Malthusian ecology, 202
"Sudden Oak Death" disease, 79
Sulfadoxine–pyrimethamine (SP), 489
Super spread event (SSE), 601
Superantigens, 560
Surface-enhanced laser desorption time of flight mass spectrometry technology (SELDI-TOF-MS technology), 235
Swine-origin influenza A virus (S-OIV), 218
Synchronized global sweep, 629–630
Synthetic insecticides, 315–317
Systematic error, 180
Systematics, 367–368
SZPE. *See S. zeamais* primary endosymbiont (SZPE)

T

T-cell epitopes, 548
TAF. *See* Tenofovir alafenamide (TAF)
Tangent space variables, 286–287
Target of rifampicin, 543–544
Target-site modification, 322–326
acetylcholinesterase, 325–326
GABA receptors, 322–323
VGSCs, 323–325
Taxa, 1, 8, 11, 42, 171
TB. *See* Tuberculosis (TB)
TBRF. *See* Tick-borne relapsing fever (TBRF)
"Tc-Bat", 520–521
TcIV, 465–466
TDF. *See* Tenofovir disoproxil fumarate (TDF)
Telomeres, 87
TEM-1 β-lactamase, 269
Temporal structure of *Anopheles gambiae*, 363
Tenofovir, 436
Tenofovir alafenamide (TAF), 435–436
Tenofovir disoproxil fumarate (TDF), 435

Downloaded From Vet-eBooks.com

Theory-based pragmatism
bacterial ecotypes cohesion, 41–42
bacterial speciation models, 27f
biological species, 26
biotechnologists, 33
cluster-based approach, 25
ecological breadth of recognized species, 30–33
ecological divergence, 26f
ecology into bacterial systematics, 43–44
ecotypes, 40–41
models of frequent speciation, 38–40
molecular technology, 32
newly divergent bacterial species, 29
putative ecotypes
ecological diversity within, 36–38
from sequence data, 35–36
species demarcations, 31f
Stable Ecotype model of bacterial speciation, 33–35
Therapeutic response, 216–217
Threonine, 613
Thrombospondin-related protein (TRAP), 490
Tick-borne relapsing fever (TBRF), 367
Ticks, 367
TLRs. *See* Toll-like receptors (TLRs)
TNF. *See* Tumor necrosis factor (TNF)
Tobacco mosaic virus, 11–12
Tobamovirus genus, 11–12
Toll-like receptors (TLRs), 198–199
Toxic shock syndrome toxin 1 (*tst*), 560
Toxin receptors, 327
Toxoplasma, 197
Toxoplasma gondii (*T. gondii*), 233–234
Trade-off hypothesis, 119
Transcriptome, 228
analysis, 428–429
Transcriptomics, 409–412, 415–416
Transferrin (TRFE), 246
Transformation, 100
Transformational evolution, 167–168
Transient mutators, 266
Transposable elements, 87
Transposition of IS6110 during infection, 544
Transthyretin (TTR), 242–246, 244f
TRAP. *See* Thrombospondin-related protein (TRAP)
Tree of Cells, 188
Tree of Life, 181, 184, 188
root, 183–184
Tree-based network, 186
TRFE. *See* Transferrin (TRFE)
Triatoma brasiliensis (*T. brasiliensis*), 292, 354
Triatoma dimidiata (*T. dimidiata*), 354, 360
Triatoma infestans (*T. infestans*), 354
Triatoma juazeirensis (*T. juazeirensis*), 292
Triatoma macromelasoma (*T. macromelasoma*), 292
Triatoma protracta (*T. protracta*), 293
Triatominae, genetics of, 351–361, 352f–353f, 355t–357t
cytogenetics, 354
evolution, 352–354
future directions, 360–361, 361f
genetic diversity, 354–360
population structure, 360
Triatominae tribes, 354
S,S,S-Tributyl phosphorotrithioate, 321
Trichinella spiralis (*T. spiralis*), 236–237
Trinity RNA-Seq *de novo* assembly software package, 415–416
Triploid hybrids, 463–464
Tritrichomonas muris (*T. muris*), 238–239
TriTryps, 459
Trypanosoma brucei (*T. brucei*), 195–196, 459
genetic crosses, 459–461
implications for epidemiology, 465
inheritance of kDNA, 464
location of genetic exchange, 461–462
Mendelian inheritance and meiosis, 462–464
Trypanosoma congolense (*T. congolense*), 238–239
Trypanosoma cruzi (*T. cruzi*), 170–171, 238–239, 351, 459, 510
behavior of experimental hybrids, 470
diversity, 465–466
genetic crosses, 467–468
genetic exchange in experimental and natural hybrids, 471–473
genome sequence of natural hybrid, 467
implications for epidemiology, 473–474
location of genetic exchange, 468–470
near-clade, 520–521

Downloaded From Vet-eBooks.com

Trypanosomatids, genetic exchange in, 459
Leishmania, 474–476
T. brucei, 459–465
T. cruzi, 465–474
Tsetse fly, 459–460, 465
breeding structures, 351
family Glossinidae, 343–344
genetics and population genetics
cytogenetics, 344
microsatellite loci and mtDNA, 344
population structure, 350
genetics of, 343–351, 343f
hemoflagellate trypanosomes, 343
trypanosomiasis control, 350–351
tsetse fly biology, 350
tsetse population management, 350
tst. *See* Toxic shock syndrome toxin 1 (*tst*)
TTR. *See* Transthyretin (TTR)
Tubercle Bacillus, host–pathogen coevolution of, 531–538
Tuberculosis (TB), 270, 444–445, 529
Tumor necrosis factor (TNF), 512–513
Turnaround time, 421, 423
Two-dimensional electrophoresis (2-DE), 229
Two-migration model, 157
Typing methods, 383–384, 387

U

UDP-glucuronosyltransferase (UGT), 437
UDP-glucuronosyltransferase 1A1 (UGT1A1), 440
UDP-glycosyl-transferases (UGTs), 319
Union of maximum parsimonious trees (UMP trees), 393
Unweighted Pair Group Method with Arithmetic Mean (UPGMA), 391

V

Vaccine, 214
development, 32
target, 214
Variable number of tandem repeats Analysis (VNTR), 150, 387
Variant surface glycoprotein (VSG), 233–234
Variational evolution, 167–168
Variograms, 83
VBN. *See* Vector Biology Network (VBN)
VDN. *See* Virology Division News (VDN)
Vector Biology Network (VBN), 217
Vector control, 217, 316–317, 341
Vector genetic diversity, 517
Vector-borne disease, 313–314, 341, 342f, 509
Vector–parasite coadaptations, 350
VGIIa/major genotype, 74–75
VGSCs. *See* Voltage-gated sodium channels (VGSCs)
Viannia subgenus, 474
Vibrio cholerae (*V. cholerae*), 60
Viral adaptation, 589–590
Viral genetic diversity generation by recombination, 612
Viral genomes, 9–11
Viral taxonomy, 1
Virology Division News (VDN), 14–15
Virtual globe software, 394
Virus species, 1–2
bionominalism, 4–5
distinction between polythetic and monothetic classes, 7f
hierarchical virus classification, 3–4
human polyomavirus, 155–159
new ICTV definition of, 11–14
non-Latinized binomial names for, 14–16
problem, 5–7
properties for, 8–9
viral genomes, 9–11
viral taxonomy, 1
Virus(es), 155–159
classification, 3–4
isolates, 13–14
Virus–host interactions, 236. *See also* Host–pathogen interactions
Visceral leishmaniasis (VL), 474
VNTR. *See* Variable number of tandem repeats Analysis (VNTR)
Voltage-gated sodium channels (VGSCs), 323–325
VSG. *See* Variant surface glycoprotein (VSG)

W

Wahlund effect, 80
West Central Africa
cross-species transmissions in HIV-1 viruses, 584
HIV-1 from SIVs circulating in chimpanzees and gorillas, 582f

Downloaded From Vet-eBooks.com

Wet market, 606
WGS. *See* Whole-genome sequencing (WGS)
White-nose syndrome, 74
WHO, 602–603
Whole-genome sequencing (WGS), 55, 414, 421, 556b–557b, 559
  analysis, 414–415
Wild bird(s)
  HPAI H5N1 and H5NX virus, 627–629
  LPAI virus
    subtypes, 624
    transmission and epidemiology, 624–626
  surveillance programs, 633–634

X
XDR TB. *See* Extensively drugresistant TB (XDR TB)

Y
Yellow fluorescent protein (YFP), 462–463
Yersinia pseudotuberculosis (*Y. pseudotuberculosis*), 37–38

Z
Zidovudine, 436
Zielh–Neelsen staining, 529
Zika virus, 408
Zoonoses, 496–498
Zygote, 459–460

Downloaded From Vet-eBooks.com

This page intentionally left blank

Downloaded From Vet-eBooks.com

GENETICS AND EVOLUTION OF INFECTIOUS DISEASES

Second Edition

Edited by: ***Michel Tibayrenc***

Maladies Infectieuses et Vecteurs Ecologie, Génétique, Evolution et Contrôle, IRD Center, Montpellier, France

Genetics and Evolution of Infectious Diseases, Second Edition discusses the constantly evolving field of infectious diseases and their continued impact on the health of populations, especially in resource-limited areas of the world. Students in public health, biomedical professionals, clinicians, public health practitioners, and decision makers will find valuable information in this book that is relevant to the control and prevention of neglected and emerging worldwide diseases that are a major cause of global morbidity, disability, and mortality.

Although substantial gains have been made in public health interventions for the treatment, prevention, and control of infectious diseases during the last century, in recent decades the world has witnessed a worldwide human immunodeficiency virus pandemic, increasing antimicrobial resistance, and the emergence of many new bacterial, fungal, parasitic, and viral pathogens.

The economic, social, and political burden of infectious diseases is most evident in developing countries, which must confront the dual burden of death and disability due to infectious illnesses.

KEY FEATURES

- Takes an integrated approach to infectious diseases
- Includes contributions from leading authorities
- Provides the latest developments in the field of infectious disease, with special focus on the most advanced technologies

Life Sciences / Microbiology

ISBN 978-0-12-799942-5

9 780127 999425

ELSEVIER

elsevier.com

Downloaded From Vet-eBooks.com