Handbook of Molecular Microbial Ecology II
Metagenomics in Different Habitats

Edited by

Frans J. de Bruijn
To my two daughters, Waverly and Vanessa de Bruijn, for their support even from a distance
Contents

Preface xiii
Contributors xv

1. Introduction 1
   Frans J. de Bruijn

Part 1 Viral Genomes

2. Viral Metagenomics 5
   Shannon J. Williamson

3. Methods in Viral Metagenomics 15
   Rebecca Vega Thurber

4. Metagenomic Contrasts of Viruses in Soil and Aquatic Environments 25
   K. Eric Wommack, Sharath Srinivasiah, Mark R. Liles, Jaysheel Bhavsar, Shellie Bench, Kurt E.
   Williamson, and Shawn W. Polson

5. Biodiversity and Biogeography of Phages in Modern Stromatolites and Thrombolites 37
   Christelle Desnues, Beltran Rodriguez-Brito, Steve Rayhawk, Scott Kelley, Taong Tran, Matthew
   Haynes, Hong Liu, Mike Furlan, Linda Wegley, Betty Chau, Yijun Ruan, Dana Hall, Florent E.
   Angly, Robert A. Edwards, Lanlin Li, Rebecca Vega Thurber, R. Pamela Reid, Janet Siefert, Valeria
   Souza, David L. Valentine, Brandon K. Swan, Mya Breitbart, and Forest Rohwer

6. Assembly of Viral Metagenomes from Yellowstone Hot Springs Reveals Phylogenetic
   Relationships and Host Co-Evolution 45
   Thomas W. Schoenfeld and David Mead
# Contents

7. Next-Generation Sequencing and Metagenomic Analysis: A Universal Diagnostic Tool in Plant Pathology 63

8. Direct Metagenomic Detection of Viral Pathogens in Human Specimens Using an Unbiased High-Throughput Sequencing Approach 73
   Takaaki Nakaya, Shota Nakamura, Yoshiko Okamoto, Yoshiyuki Nagai, Jun Kawai, Yoshihide Hayashizaki, Tetsuya Iida, and Toshihiro Horii

## Part 2 The Soil Habitat

9. Soil-Based Metagenomics 83
   Rolf Daniel

10. Methods in Metagenomic DNA, RNA, and Protein Isolation from Soil 93
    T. Rajesh, J. Rajendhran, P. Lavanya Pushpam, and P. Gunasekaran

11. Soil Microbial DNA Purification Strategies for Multiple Metagenomic Applications 109
    Larissa C. Parsley, Chengcang Wu, David Mead, Robert M. Goodman, and Mark R. Liles

    Sho Morimoto and Takeshi Fujii

13. Actinobacterial Diversity Associated with Antarctic Dry Valley Mineral Soils 125
    Bronwyn M. Kirby, Marilize Le Roes-Hill, S. Craig Cary, Stephanie G. Burton, I. Marla Tuffin, and Don A. Cowan

14. Targeting Major Soil-Borne Bacterial Lineages Using Large-Insert Metagenomic Approaches 135
    Anna M. Kielak and George A. Kowalchuk

15. Novelty and Uniqueness Patterns of Rare Members of the Soil Biosphere 143
    Mostafa S. Elshahed and Noha H. Youssef

16. Extensive Phylogenetic Analysis of a Soil Bacterial Community Illustrates Extreme Taxon Evenness and the Effects of Amplicon Length, Degree of Coverage, and DNA Fractionation on Classification and Ecological Parameters 151
    Sergio E. Morales, Theodore F. Cosart, Jesse V. Johnson, and William E. Holben

17. The Antibiotic Resistome: Origins, Diversity, and Future Prospects 165
    Erin L. Westman and Gerard D. Wright
Contents

Part 3  The Digestive Tract

18. Functional Intestinal Metagenomics 177
   Bartholomeus van den Bogert, Milkha M. Leimena, Willem M. de Vos, Erwin G. Zoetendal, and Michiel Kleerebezem

19. Assessment and Improvement of Methods for Microbial DNA Preparation from Fecal Samples 191
   Mariko Ueno, Mami Kikuchi, Kenshiro Oshima, Seok-won Kim, Hidetoshi Morita, and Masahira Hattori

20. The Role of Dysbiosis in Inflammatory Bowel Diseases 199
   Johan Dicksved and Ben Willing

   Jonathan Swann, Selena E. Richards, Qing Shen, Elaine Holmes, Julian R. Marchesi, and Kieran Tuohy

22. Complete Genome of an Uncultured Endosymbiont Coupling Nitrogen Fixation to Cellulolysis within Protist Cells in Termite Gut 221
   Yuichi Hongoh

23. Cloning and Identification of Genes Encoding Acidic Cellulases from the Metagenomes of Buffalo Rumen 229
   Cheng-Jie Duan, Jun-Liang Liu, and Jia-Xun Feng

Part 4  Marines and Lakes

24. Microbial Diversity in the Deep Sea and the Underexplored “Rare Biosphere” 245
   David B. Mark Welch and Susan M. Huse

25. Bacterial Community Structure and Dynamics in a Seasonally Anoxic Fjord: Saanich Inlet, British Columbia 253
   David A. Walsh and Steven J. Hallam

26. Adaptation to Nutrient Availability in Marine Microorganisms by Gene Gain and Loss 269
   Adam C. Mariiny, Ying Huang, and Weizhong Li

27. Detection of Large Numbers of Novel Sequences in the Metatranscriptomes of Complex Marine Microbial Communities 277
   Jack A. Gilbert, Dawn Field, Ying Huang, Robert A. Edwards, Weizhong Li, Paul Gilna, and Ian Joint
Contents

   Didier Debroas, François Enault, Isabelle Jouan-Dufournel, Gisèle Bronner, and Jean-François Humbert

29. Metagenomics of the Marine Subsurface: The First Glimpse from the Peru Margin, ODP Site 1229 295
   Jennifer F. Biddle

30. A Targeted Metagenomic Approach to Determine the “Population Genome” of Marine Synechococcus 301
   Sophie Mazard, Martin Ostrowski, Laurence Garczarek, and David J. Scanlan

31. Diversity and Role of Bacterial Integron/Gene Cassette Metagenome in Extreme Marine Environments 309
   Hosam Elsaied and Akihiko Maruyama

Part 5  Other Habitats

32. The Olavius algarvensis Metagenome Revisited: Lessons Learned from the Analysis of the Low-Diversity Microbial Consortium of a Gutless Marine Worm 321
   Manuel Kleiner, Tanja Woyke, Caroline Ruehland, and Nicole Dubilier

33. Microbiome Diversity in Human Saliva 335
   Ivan Nasidze and Mark Stoneking

34. Approaches to Understanding Population Level Functional Diversity in a Microbial Community 341
   Devaki Bhaya

35. A Functional Metagenomic Approach for Discovering Nickel Resistance Genes from the Rhizosphere of an Acid Mine Drainage Environment 355
   Salvador Mirete, Carolina G. de Figueras, and Jose E. Gonzalez-Pastor

36. The Microbiome of Leaf-Cutter Ant Fungus Gardens 367
   Garret Suen, Jarrod J. Scott, Frank O. Ayward, and Cameron R. Currie

37. Diversity of Archaea in Terrestrial Hot Springs and Role in Ammonia Oxidation 381
   Chuanlun L. Zhang, Brian P. Hedlund, and Jun Meng

38. Colonization of Nascent, Deep-Sea Hydrothermal Vents by a Novel Archaeal and Nanoarchaeal Assemblage 395
   Thomas D. Niederberger, Elizabeth A. McCliment, and S. Craig Cary
39. Analysis of the Metagenome from a Biogas-Producing Microbial Community by Means of Bioinformatics Methods 403
  Sebastian Jaenicke, Martha Zakrzewski, Sebastian Junemann, Alfred Pühler, Alexander Goesmann, and Andreas Schlüter

40. Amplicon Pyrosequencing Analysis of Endosymbiont Population Structure 415
  Frank J. Stewart and Colleen M. Cavanaugh

41. Investigating Bacterial Diversity Along Alkaline Hot-Spring Thermal Gradients by Barcoded Pyrosequencing 423
  Scott R. Miller and Michael Weltzer

42. Genetic Characterization of Microbial Communities Living at the Surface of Building Stones 429
  Maité Berdoulay and Jean-Claude Salvado

Part 6 Biodegradation

43. Novel Aromatic Degradation Pathway Genes and their Organization as Revealed by Metagenomic Analysis 439
  Kentaro Miyazaki

44. Functional Screening a Wide Host-Range Metagenomic Library from a Wastewater Treatment Plant Yields a Novel Alcohol/Aldehyde Dehydrogenase 451
  Margaret Wexler, Philip L. Bond, David J. Richardson, and Andrew W. B. Johnston

45. Aromatic Hydrocarbon Degradation Genes from Chronically Polluted Subantarctic Marine Sediments 461
  Hebe M. Dionisi, Mariana Lozada, Magali S. Marcos, Walter D. Di Marzio, and Claudia L. Loviso

46. Isolation and Characterization of Alkane Hydroxylases from a Metagenomic Library of Pacific Deep-Sea Sediment 475
  Fengping Wang, Meixiang Xu, and Xiang Xiao

Part 7 Biocatalysts and Natural Products

47. Emerging Fields in Functional Metagenomics and Its Industrial Relevance: Overcoming Limitations and Redirecting the Search for Novel Biocatalysts 483
  Mirjam Perner, Nele Imberger, Hans Ulrich Köhler, Jennifer Chow, and Wolfgang R. Streit

48. Carboxylesterases and Lipases from Metagenomes 499
  Jennifer Chow, Ulrich Krauss, Karl-Erich Jaeger, and Wolfgang R. Streit
49. Expanding Small-Molecule Functional Metagenomics through Parallel Screening of Broad Host-Range Cosmid Environmental DNA Libraries in Diverse Proteobacteria 507
   Jeffrey W. Craig and Sean F. Brady

50. Biomedicinals from the Microbial Metagenomes of Marine Invertebrates 517
   Walter C. Dunlap, Paul F. Long, and Marcel Jaspars

51. Molecular Characterization of TEM-Type Beta-Lactamases Identified in Cold-Sea Sediments of Edison Seamount (South of Lihir Island, Papua New Guinea) 545
   Sang Hee Lee and Jung-Hyun Lee

52. Identification of Novel Bioactive Compounds from the Metagenome of the Marine Sponge Haliclona simulans 553
   David P. H. Lejon, Jonathan Kennedy, and Alan D. W. Dobson

53. Functional Viral Metagenomics and the Development of New Enzymes for DNA and RNA Amplification and Sequencing 563
   Thomas W. Schoenfeld, Nick Hermersmann, Mike Moser, Darby Renneckar, Vinay Dhodda, and David Mead

Part 8 Summary

54. Metagenomics: The Paths Forward 581
   C. Titus Brown and James M. Tiedje

55. Darwin in the Twenty-First Century: Natural Selection, Molecular Biology, and Species Concepts 589
   Francisco J. Ayala

Index 597
Preface

In the last 25 years, microbiology and molecular microbial ecology have undergone drastic transformations that changed the microbiologist's view of how to study microorganisms. The main problem before was the assumption that microorganisms needed to be culturable, in order to classify them and study their metabolic and organismal diversity. The heart of this transformation was the convincing demonstration that the yet unculturable world was far greater than the culturable one. In fact, the number of microbial genomes has been estimated to be between 2000 to 18,000 genomes per gram of soil. In 1985, an experimental advance radically changed our perception of the microbial world. After Carl Woese showed that rRNA genes could be used to derive evolutionary relationships, phylogenetic “trees” and evolutionary chromometers, Norman Pace and colleagues created a new chapter in molecular microbial ecology, using the direct analysis of rRNA sequences in the environment to describe the diversity of microorganisms without culturing [Handelsman, 2004]. The next major step forward was the development of the PCR reaction, to amplify rRNA genes for subsequent sequence analysis and classification. The subsequent major advance was the notion that one could extract total DNA or RNA from environmental samples, including culturable and yet unculturable organisms, and clone it into a suitable vector for introduction into a culturable organism, followed by analysis by using high throughput shotgun DNA sequencing of cloned DNA, or by direct sequencing. The idea of cloning DNA directly from environmental samples was first proposed by Page; this method was coined “metagenomics” by Handelsman et al. in 1994, and is now used in many laboratories worldwide to study diversity and for the isolation of novel medical and industrial compounds.

These recent studies are reviewed in this book and the companion book, *Handbook of Molecular Microbial Ecology I: Metagenomics and Complementary Approaches*. Instead of relying only on a limited number of (long) review articles on selected topics, Volume I provides reviews as well as a large number of case studies, mostly based on original publications and written by expert “at-the-bench” scientists from more than 20 different countries. These books highlight the databases and computer programs used in each study, by listing them at the end of the chapter, together with their sites. This is a special feature of both books, facilitating the computer-assisted analysis of the vast amount of data generated by metagenomic studies.

In addition, this book describes metagenomic studies in a variety of habitats which present a large number of system dependent different approaches in greatly differing habitats. These approaches also result in the presentation of multiple biological systems which are interesting to microbial ecologists and microbiologists in their own right. Both books should be of interest to scientists in the fields of soil, water, medicine and industry who are or are contemplating using metagenomics and complementary approaches to address academic, medical, or industrial questions about bacterial communities from varied habitats, but also to those interested in particular biological systems in general.

ACKNOWLEDGMENTS

For their support of this project, I gratefully acknowledge: The Laboratory for Plant Microbe Interactions (LIPM), the Institut National de Recherche de Agriculture (INRA), and the Centre National de Recherche Scientifique (CNRS).

I would like to thank Claude Bruand for his help with the computer work.

Faans J. de Brujin

Castanet, Tolosan, France

May 2011
Contributors

Editor
Frans J. de Bruijn, Laboratory of Plant Micro-organism Interaction, CNRS-INRA, Castanet Tolosan, France

Authors
Ian P. Adams, Food and Environment Research Agency, York, United Kingdom
Florent E. Angly, San Diego State University, San Diego, California
Francisco J. Ayala, University of California, Irvine, California
Frank O. Aylward, University of Wisconsin, Madison, Wisconsin
Shellie Bench, Ocean Sciences Department, University of California, Santa Cruz, California
Maite Berdoulay, University of Pau and Countries of the Adour, Anglet, France; Nobatek, Anglet, France
Jaysheel Bhavsar, University of Maryland School of Medicine, Baltimore, Maryland
Devaki Bhaya, Carnegie Institution for Science, Stanford, California
Jennifer F. Biddle, University of Delaware, Lewes, Delaware
Philip L. Bond, University of East Anglia, Norwich, United Kingdom
Neil Boonham, Food and Environment Research Agency, York, United Kingdom
Sean F. Brady, The Rockefeller University, New York, New York
Mya Breitbart, University of South Florida, St. Petersburg, Florida
Gisèle Bronner, Clermont University, University of Blaise Pascal, Paris, France
C. Titus Brown, Michigan State University, East Lansing, Michigan
Stephanie G. Burton, Cape Peninsula University of Technology, Bellville, South Africa
S. Craig Cary, University of Delaware, Lewes, Delaware; University of Waikato, Hamilton, New Zealand
Colleen M. Cavanaugh, Harvard University, Cambridge, Massachusetts
Betty Chau, San Diego State University, San Diego, California
Jennifer Chow, University of Hamburg, Hamburg, Germany
Theodore F. Cosart, The University of Montana, Missoula, Montana
Don A. Cowan, University of the Western Cape, Cape Town, South Africa
Jeffrey W. Craig, The Rockefeller University, New York, New York
Cameron R. Currie, University of Wisconsin, Madison, Wisconsin
Rolf Daniel, Georg August University of Göttingen, Göttingen, Germany
Carolina G. de Figueras, Laboratory of Molecular Ecology, Center for Astrobiology (CSIC-INTA), Madrid, Spain
Contributors

Willem M. de Vos, Wageningen University, Wageningen, The Netherlands; University of Helsinki, Helsinki, Finland
Didier Debroas, Clermont University, University of Blaise Pascal, Paris, France
Christelle Desnues, San Diego State University, San Diego, California
Vinay Dhodda, Lucigen Corporation, Middleton, Wisconsin
Walter D. Di Marzio, National University of Luján–CONICET, Buenos Aires, Argentina
Johan Dickved, Swedish University of Agricultural Sciences, Uppsala, Sweden
Hebe M. Dionisi, Environmental Microbiology Laboratory, Patagonian National Research Center (CENPAT–CONICET), Puerto Madryn, Chubut Province, Argentina
Alan D. W. Dobson, University College Cork, Cork, Ireland
Cheng-Jie Duan, Guangxi University, Guangxi, China
Nicole Dubilier, Max Planck Institute for Marine Microbiology, Bremen, Germany
Walter C. Dunlap, Australian Institute of Marine Science, Queensland, Australia
Robert A. Edwards, San Diego State University, San Diego, California; Argonne National Laboratory, Argonne, Illinois
Hosam Elsaied, National Institute of Advanced Industrial Science and Technology, Ibaraki, Japan
Mostafa S. Elshahed, Oklahoma State University, Stillwater, Oklahoma
François Enault, Clermont University, University of Blaise Pascal, Paris, France
Jia-Xun Feng, Guangxi University, Guangxi, China
Dawn Field, NERC Centre for Ecology and Hydrology, Oxford, United Kingdom
Takeshi Fujii, National Institute for Agro-Environmental Sciences, Ibaraki, Japan
Mike Furlan, San Diego State University, San Diego, California
Laurence Garzarek, University of Paris 6 and CNRS, Roscoff, France
Jack A. Gilbert, Plymouth Marine Laboratory, Plymouth, United Kingdom
Paul Gilna, University of California, San Diego, La Jolla, California
Rachel H. Glover, Food and Environment Research Agency, York, United Kingdom
Alexander Goessmann, Bielefeld University, Bielefeld, Germany
José E. González-Pastor, Laboratory of Molecular Ecology, Center for Astrobiology (CSIC-INTA), Madrid, Spain
Robert M. Goodman, Rutgers, The State University of New Jersey, New Brunswick, New Jersey
P. Gunasekaran, Madurai Kamaraj University, Madurai, India
Dana Hall, San Diego State University, San Diego, California
Steven J. Hallam, University of British Columbia, Vancouver, British Columbia, Canada
Masahira Hattori, The University of Tokyo, Chiba, Japan
Matthew Haynes, San Diego State University, San Diego, California
Yoshihide Hayashizaki, Omics Science Center (OSC), RIKEN, Kanagawa, Japan
Brian P. Hedlund, University of Nevada, Las Vegas, Nevada
Nick Hermsmann, Lucigen Corporation, Middleton, Wisconsin
William E. Holben, The University of Montana, Missoula, Montana
Elaine Holmes, Imperial College London, London, United Kingdom
Yuichi Hongoh, Tokyo Institute of Technology, Tokyo, Japan
Toshihiro Hori, Osaka University, Suita, Osaka, Japan
Ying Huang, University of California, San Diego, La Jolla, California
Jean-François Humbert, Clermont University, University of Blaise Pascal, Paris, France
Susan M. Huse, Marine Biological Laboratory at Woods Hole, Woods Hole, Massachusetts
Contributors

Tetsuya Iida, Osaka University, Suita, Osaka, Japan
Nele Ilmberger, University of Hamburg, Hamburg, Germany
Elena Jackeviciene, Plant Protection Service of Lithuania, Vilnius, Lithuania
Karl-Erich Jaeger, Heinrich-Heine-University Duesseldorf, Juelich, Germany
Sebastian Juenicke, Bielefeld University, Bielefeld, Germany
Marcel Jaspars, University of Aberdeen, Scotland, United Kingdom
Jesse V. Johnson, The University of Montana, Missoula, Montana
Andrew W. B. Johnston, University of East Anglia, Norwich, United Kingdom
Ian Joint, Plymouth Marine Laboratory, Plymouth, United Kingdom
Isabelle Jouan-Dufournel, Clermont University, University of Blaise Pascal, Paris, France
Sebastian Juenemann, Bielefeld University, Bielefeld, Germany
Jun Kawai, Omics Science Center (OSC), RIKEN, Kanagawa, Japan
Anna M. Kielak, Netherlands Institute of Ecology, Heteren, The Netherlands; University of Groningen, The Netherlands
Scott Kelley, San Diego State University, San Diego, California
Jonathan Kennedy, University College, Cork, Ireland
Mani Kikuchi, Kureha Corporation, Tokyo, Japan; The University of Tokyo, Chiba, Japan
Seok-won Kim, The University of Tokyo, Chiba, Japan
Bronwyn M. Kirby, University of the Western Cape, Cape Town, South Africa
Michiel Kleerebezem, Top Institute Food and Nutrition, Wageningen, The Netherlands; Wageningen University, Wageningen, The Netherlands
Manuel Kleiner, Max Planck Institute for Marine Microbiology, Bremen, Germany
Hans Ulrich Kohler, University of Hamburg, Hamburg, Germany
George A. Kowalchuk, Netherlands Institute of Ecology, Heteren, The Netherlands; Free University Amsterdam, Amsterdam, The Netherlands
Ulrich Krauss, Heinrich Heine University Duesseldorf, Juelich, Germany
Jung-Hyun Lee, Korea Ocean Research & Development Institute, Ansan, Republic of Korea
Sang Hee Lee, Myongji University, Seoul, Republic of Korea
Milka M. Leimena, Top Institute Food and Nutrition, Wageningen, The Netherlands; Wageningen University, Wageningen, The Netherlands
David P. H. Lejon, University College, Cork, Ireland
Linlin Li, San Diego State University, San Diego, California
Weizhong Li, University of California, San Diego, La Jolla, California
Mark R. Liles, Auburn University, Auburn, Alabama
Jun-Liang Liu, Guangxi University, Guangxi, China
Hong Liu, San Diego State University, San Diego, California
Paul F. Long, University of London, London, United Kingdom
Claudia L. Loviso, Environmental Microbiology Laboratory, Patagonian National Research Center (CENPAT—CONICET), Puerto Madryn, Chubut Province, Argentina
Mariana Lozada, Environmental Microbiology Laboratory, Patagonian National Research Center (CENPAT—CONICET), Puerto Madryn, Chubut Province, Argentina
Julian Marchesi, Cardiff University, Cardiff, Scotland, United Kingdom
Magali S. Marcos, Environmental Microbiology Laboratory, Patagonian National Research Center (CENPAT—CONICET), Puerto Madryn, Chubut Province, Argentina
Contributors

Adam C. Martiny, University of California, Irvine, California
Akihiko Maruyama, National Institute of Advanced Industrial Science and Technology, Ibaraki, Japan
Sophie Mazzard, University of Warwick, Coventry, United Kingdom
Elizabeth A. McCliment, Marine Biological Laboratory, Woods Hole, Massachusetts,
David Mead, Lucigen Corporation, Middleton, Wisconsin
Jun Meng, Xiamen University, Xiamen, China
Scott R. Miller, University of Montana, Missoula, Montana
Salvador Mirete, Laboratory of Molecular Ecology, Center for Astrobiology (CSIC-INTA), Madrid, Spain
Kenzo Miyazaki, National Institute of Advanced Industrial Science and Technology (AIST), Ibaraki, Japan; The University of Tokyo, Ibaraki, Japan
Wendy A. Monger, Food and Environment Research Agency, York, United Kingdom
Sergio E. Morales, The University of Montana, Missoula, Montana
Sho Morimoto, National Institute for Agro-Environmental Sciences, Ibaraki, Japan
Hidetoshi Morita, Azabu University, Kanagawa, Japan
Mike Moser, Lucigen Corporation, Middleton, Wisconsin
Rick Mumford, Food and Environment Research Agency, York, United Kingdom
Yoshikazu Nagai, Center of Research Network for Infectious Diseases (CRNID), RIKEN, Tokyo, Japan
Shota Nakamura, Osaka University, Suita, Osaka, Japan
Takanori Nakaya, Osaka University, Suita, Osaka, Japan
Ivan Nadtich, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
Meletele Navalaskiene, Plant Virus Laboratory, Institute of Botany, Vilnius, Lithuania
Thomas D. Niederberger, University of Delaware, Lewes, Delaware
Yoshiko Okamoto, Center of Research Network for Infectious Diseases (CRNID), RIKEN, Tokyo, Japan
Kenshiro Oshimura, The University of Tokyo, Chiba, Japan
Martin Ostrowski, University of Warwick, Coventry, United Kingdom
Larissa C. Parsley, Auburn University, Auburn, Alabama
Mirjam Perner, University of Hamburg, Hamburg, Germany
Shawn W. Polson, Delaware Biotechnology Institute, Newark, Delaware
Alfred Pühler, Bielefeld University, Bielefeld, Germany
P. Lavanya Pushpan, Madurai Kamaraj University, Madurai, India
R. Rajendran, Madurai Kamaraj University, Madurai, India
T. Rajesh, Madurai Kamaraj University, Madurai, India
Steve Rayhawk, San Diego State University, San Diego, California
R. Pamela Reid, University of Miami, Miami, Florida
Darby Renneckar, Lucigen Corporation, Middleton, Wisconsin
Selena E. Richards, Imperial College London, London, United Kingdom
David J. Richardson, University of East Anglia, Norwich, United Kingdom
Beltran Rodriguez-Brito, San Diego State University, San Diego, California
Yijun Ruan, Genome Institute of Singapore, Singapore
Caroline Ruechland, Max Planck Institute for Marine Microbiology, Bremen, Germany
Jean-Claude Salvado, University of Pau and Countries of the Adour, Anglet, France
Contributors

Marija Samuitiene. Plant Virus Laboratory, Institute of Botany, Vilnius, Lithuania
David J Scanlan. University of Warwick, Coventry, United Kingdom
Andreas Schlüter. Bielefeld University, Bielefeld, Germany
Thomas W. Schoenfeld. Lucigen Corporation, Middleton, Wisconsin
Jarrod J. Scott. University of Wisconsin, Madison, Wisconsin
Qing Shen. The University of Reading, Whiteknights, Reading, United Kingdom
Janet Siefert. Rice University, Houston, TX
Valeria Souza. Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico
Sharath Srinivasiah. Delaware Biotechnology Institute, Newark, Delaware
Frank J. Stewart. Massachusetts Institute of Technology, Cambridge, Massachusetts
Mark Stoneking. Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
Wolfgang R. Streit. University of Hamburg, Hamburg, Germany
Garret Suen. University of Wisconsin, Madison, Wisconsin
Brandon K. Swan. University of California at Santa Barbara, Santa Barbara, California
Jonathan Swann. Imperial College London, London, United Kingdom
Rebecca Vega Thurber. Florida International University, North Miami, Florida; San Diego State University, San Diego, California
Richard Thwaites. Food and Environment Research Agency, York, United Kingdom
James M. Tiedje. Michigan State University, East Lansing, Michigan
Tuong Tran. San Diego State University, San Diego, California
I. Marla Tuffin. University of the Western Cape, Cape Town, South Africa
Kieran Tuohy. IASMA Research and Innovation Center, Michele, Italy
Mariko Ueno. Kureha Corporation, Tokyo, Japan
David L. Valentine. University of California, Santa Barbara, California
David A. Walsh. University of British Columbia, Vancouver, British Columbia, Canada
Fengping Wang. Key Laboratory of Marine Biogenetic Resources, State Oceanic Administration, Xiamen, China; Shanghai JiaoTong University, Shanghai, China
Linda Wegley. San Diego State University, San Diego, California
David B. Mark Welch. Marine Biological Laboratory at Woods Hole, Woods Hole, Massachusetts
Michael Weltzer. University of Montana, Missoula, Montana
Erin L. Westman. McMaster University, Hamilton, Ontario, Canada
Margaret Westler. University of East Anglia, Norwich, United Kingdom
Kurt E. Williamson. The College of William and Mary, Williamsburg, Virginia
Shannon J. Williamson. J. Craig Venter Institute, San Diego, California
K. Eric Wommack. Delaware Biotechnology Institute, Newark, Delaware
Tanja Woyke. Department of Energy Joint Genome Institute, Walnut Creek, California
Gerard D. Wright. McMaster University, Hamilton, Ontario, Canada
Bartholomeus van den Bogert. Top Institute Food and Nutrition, Wageningen, The Netherlands; Wageningen University, Wageningen, The Netherlands
Chenguang Wu. Lucigen Corporation, Middleton, Wisconsin
Xiang Xiao. Key Laboratory of Marine Biogenetic Resources, State Oceanic Administration, Xiamen, China; Shanghai JiaoTong University, Shanghai, China
Meixiang Xu. Key Laboratory of Marine Biogenetic Resources, State Oceanic Administration, Xiamen, China
Contributors

Noha H. Youssef, Oklahoma State University, Stillwater, Oklahoma
Martha Zakrzewski, Bielefeld University, Bielefeld, Germany
Chuanlun L. Zhang, University of Georgia, Athens, Georgia; Tongji University, Shanghai China
Erwin G. Zoetendal, Top Institute Food and Nutrition, Wageningen, The Netherlands; Wageningen University,
Wageningen, The Netherlands
Chapter 1

Introduction

FRANS J. DE BRUIJN

In this second volume of *Handbook of Molecular Microbial Ecology*, examples are given of metagenomic studies in a large variety of habitats and using diverse techniques. Part 1 discusses the Metagenomics of Viral Genomes, with an Introduction in Chapter 2 and sample Chapters on viruses in soil and aquatic environments, modern strombolites and thrombolites, Yellowstone hot springs, human specimens and plants. In the last case, Chapter 7 proposes next-generation sequencing and metagenomic analysis as a novel, universal diagnostic tool in plant virology. Various methods are described to generate viral metagenomes and to assemble and analyze them.

In Part 2, the soil habitat is the subject, with an Introduction in Chapter 9 and topics addressed include methods for soil DNA and RNA isolation and purification for multiple metagenomics applications (Chapters 10 and 11). These techniques are essential for construction of (large insert) clone libraries (see also Volume I, Chapter 22) and random DNA sequencing studies. New approaches to retrieve full-length functional genes and phylogenetic analysis of bacterial populations or major soil-borne lineages, as well as rare members of the soil biosphere (Chapter 15), using the 16S rRNA gene and metagenomic libraries, are presented. The soil antibiotic resistome is also analyzed (Chapter 17).

In Part 3, chapters on the digestive tract are presented with an Introduction in Chapter 18, which includes the human gut microbiota, and the possible correlation of human disease with human microbiota (Chapters 18–21). Some of these studies are part of consortia such as The Human Microbiome Project, and The Human Gutna Microbiome Initiative, which are discussed in Volume I, Chapter 35. This part is complemented with chapters addressing the metagenomics of termite guts and buffalo rumens (Chapters 22 and 23).

In Part 4, the metagenomics of microbiota in the marine and lake habitats are the subject of study. Microbial diversity in the deep sea, deep sediments and the underexplored “rare biosphere,” as well as lakes is investigated (Chapters 24–31). Genomic adaptations in marine organisms (Chapter 26), the ecological genomics of marine Pico/cyanobacteria (Chapter 30) and the diversity and role of bacterial integron/gene cassette metagenomes in extreme environments (Chapter 31; see also Volume I, Chapter 26) are highlighted. These studies are complemented by a metatranscriptome analysis of complex marine microbial communities (Chapter 27; see also Volume I, Chapters 62–64).

In Part 5, metagenomic analysis of microbes in a varied number of habitats is presented, ranging from gutless marine worms, human saliva, an acid mine draining environment, terrestrial hot springs, deep-sea hydrothermal vents, biogas plants, visicomyid host clams, to the surface of building stones (Chapters 32–42). The purpose of this section is to expose the reader to many different habitats and metagenomic approaches to study them.

In Part 6, studies on the application of metagenomics to the discovery of biodegradation genes/enzymes from different habitats, such as aromatic degradation pathway genes, benzoate degradation genes, an alcohol/aldehyde dehydrogenase gene, and alkane hydroxylase genes are presented (Chapters 43–46).

In Part 7, the metagenomic discovery of several novel natural products by different methods is presented. An
overview of “functional Metagenomics and its industrial relevance” is presented in Chapter 47. The examples shown include a cold active lipase, coenzyme B12 dependent glycerol dehydratase- and diol dehydrogenases, biomedicals, and antibiotics, as well as the discovery, development and commercialization of Pyrophage 3173 DNA Polymerase (Chapters 48–53).

These parts do not mean to be all inclusive, but should serve the reader with examples of different approaches to use in their own systems/habitats, as well as provide references back to the original publication(s) the chapter was derived from and an extensive literature on the topic.

Volume II ends with a summary section comprising a perspective on the future of the “omics” and single-cell analysis (Chapter 54), as well as an article on the birthday of Darwin’s “On the Origin of Species” and the relevance of Darwin’s work to today’s molecular methods and species concepts (Chapter 55).
Part 1

Viral Genomes
Chapter 2

Viral Metagenomics

SHANNON J. WILLIAMSON

2.1 INTRODUCTION

The term “metagenomics” was coined by the soil microbial ecologist Dr. Jo Handelsman in 1998 [Handelsman et al., 1998]. Metagenomics, or community genomics, refers to the study of the genomic contents of microbes extracted directly from the environment. The establishment of metagenomic techniques was an important breakthrough in microbial ecology because microbes that can be cultivated in the laboratory are thought to account for less than 1% that exist in many environments [Whitman et al., 1998]. Over the past decade, metagenomics has developed into an emerging field of study for researchers specializing in diverse disciplines, and metagenomes have been created from the simplest biological complexes (i.e., viruses—the subject of this review) as well as from assemblages of eukaryotes. Metagenomic sequence data are typically used to address the following two fundamental questions: Who is there? and What are they doing? The taxonomic and functional data from metagenomic studies have revolutionized our understanding of the diversity of microbes and the roles they play in their communities.

Viruses are abundant and ubiquitous biological components of every biome on Earth and outnumber all cellular forms of life. Viruses have been the subject of scrutiny for nearly 120 years, but the field of viral metagenomics is relatively young, with the first marine viral metagenome published in 2002 [Breitbart et al., 2002]. Since this time, the number of published viral metagenomes (viromes) has exploded, and the adoption of next-generation sequencing technologies by researchers has resulted in a relative deluge of information on the genomic contents of viral communities inhabiting a diverse range of environments (see Chapter 3 in this volume for a comprehensive list of viromes). Analysis of viromes has enabled a deeper understanding of virus community dynamics including genotypic and taxonomic diversity, functional capacity, biogeography, and evolution. Viral metagenomics is also a powerful tool for viral discovery and has been used in this capacity to reveal the presence of novel DNA and RNA-containing viruses in a variety of samples, ranging from seawater to domesticated plants. This chapter reviews the technical and applied aspects of viral metagenomics, highlighting examples from both “natural” and human-derived environments.

2.2 EXPERIMENTAL APPROACHES AND SEQUENCING TECHNOLOGIES

Due to their ubiquitous nature, it’s possible to collect viruses from almost all types of biological samples. Indeed, viral metagenomic approaches have been applied to samples collected from a diverse range of environments: from marine waters and sediments to the human gut to a vineyard [Breitbart et al., 2002, 2003, 2004a, 2008; Angly et al., 2006; Fierer et al., 2007; Desnues et al., 2008; see also Chapter 5, Vol. II; Kim et al., 2008; McDaniel et al., 2008; Schoenfeld et al., 2008; see also Chapter 6, Vol. II, Vega Thurber et al., 2008; Williamson et al., 2008; Djikeng et al., 2009; Nakamura et al., 2009; Coetzee et al., 2010]. However, the techniques used to extract virus particles vary and are dependent on the type of sample under study. While the isolation and concentration of viruses from aquatic ecosystems is rather straightforward, more complex matrices (such as soils, sediments, tissues, and clinical samples) present a greater challenge. Collection and purification of virus particles is followed by the targeted extraction of viral nucleic acids, either DNA or RNA, with nucleic-mediated destruction...
of nontargeted molecules. Alternatively, all types of viral nucleic acids (double-stranded or single-stranded DNA or total RNA) can be purified from a sample simultaneously using hydroxyapatite chromatography [Andrews-Pfannkoch et al., 2010].

Amplification of viral DNA is often performed in order to (1) provide sufficient quantities of nucleic acid for library construction and sequencing, (2) produce unmodified copies of viral DNA [Breitbart et al., 2002], and (3) purify the DNA of potential contaminants that may interfere with downstream molecular applications [Thurber et al., 2009]. Amplification can be accomplished using a linker-mediated approach [Andrews-Pfannkoch et al., 2010] or by multiple displacement amplification using the phi29 DNA polymerase [Thurber et al., 2009]. Lastly, depending on the sequencing technology to be employed, clone-dependent or clone-independent libraries will be constructed in preparation of sequencing. For a more thorough description of the protocols used for generating viromes, see Chapter 3 in this volume.

There are currently four options when selecting a sequencing platform for metagenomic studies including di-deoxy sequencing (Sanger), pyrosequencing (454-Roche), SOLiD™ (Applied Biosystems) and Illumina® (formerly known as Solexa). Each technology has pros and cons with respect to sequencing performance including overall cost, read length, error rates, and total capacity (see Chapter 18, Vol. I). To date, only Sanger and 454 pyrosequencing have been utilized in viral metagenomic studies. Sanger sequencing, the only option available when the first viral metagenomic study was undertaken [Breitbart et al., 2002], still affords the longest read lengths of all available sequencing technologies [Wommack et al., 2008]. However, the sheer volume of data, increasing read lengths, and cost advantage afforded by pyrosequencing has resulted in a sharp decline in Sanger-based metagenomic projects in the past several years.

### 2.3 ENVIRONMENTAL STUDIES

The diversity of natural ecosystems on our planet presents unparalleled opportunities for viral metagenomic studies, and a tremendous amount of data on virus communities inhabiting a multitude of environments has been produced over a relatively short time span. The majority of viral metagenomic studies have focused on dsDNA-containing viruses, although targeted studies of viruses with alternate nucleic acid types are increasing [Culley et al., 2006; Ng et al., 2009a,b; Andrews-Pfannkoch et al., 2010]. Due to the extensive nature of environmental viral metagenomic studies, it’s prohibitive to discuss them all in detail. Therefore, this part of the chapter will highlight the significant observations generated from studies conducted on samples collected from aquatic, terrestrial, and extreme ecosystems. Figure 2.1 shows the distribution of environments from which viral metagenomes have been produced. Most studies have focused on the marine environment, although numerous hypersaline viromes have also been created. A viral metagenome has even been generated from a vineyard, effectively establishing a connection between the disciplines of viral ecology, plant pathology, and oenology [Coetzee et al., 2010].

Despite the disparate physical and chemical factors that characterize the environments shown in Figure 2.1, the resultant viromes generally share the following three characteristics: (1) a high incidence of unknown sequences, (2) a high level of genotypic diversity, and (3) evidence of functional and metabolic plasticity. The propensity of unknown sequences, or sequences that share no significant similarity to other sequences in public databases, suggests that environmental viruses are the most uncharacterized and genetically novel biological components on our planet. The high number of estimated virus genotypes in several environments undoubtedly contributes to the novelty of viral metagenomic data as well as substantial evolutionary divergence from viruses within public databases (see Chapter 4 in this Volume for additional information). Functional profiling of environmental metagenomes has perhaps revealed the most intriguing observations with respect to virus–host dynam-ics, adaptation, and evolution. A multitude of studies have now demonstrated that the adoption of environmentally relevant host genes by viruses, predominantly phage, is a common occurrence (see Rohwer and Thurber [2009] for a review of this topic; also see Dinsdale et al. [2008] and Williamson et al. [2008a]). Expression of host-derived genes is hypothesized to increase viral fitness by prolonging the life of the host while increasing replication efficiency. This phenomenon also has implications for
2.3 Environmental Studies

2.3.1 Aquatic Environments: Marine

Viral metagenomics has its roots in the marine environment, with the first study focusing on viral communities collected from two bodies of water off of Southern California [Breitbart et al., 2002]. Since this time, many viromes have been created from marine-related material including planktonic samples [Breitbart et al., 2002, 2004; Angly et al., 2006; Culley et al., 2006; Bench et al., 2007; Sharon et al., 2007, 2009; McDaniel et al., 2008; Williamson et al., 2008a] and marine sediments [Breitbart et al., 2004] and marine animals [Vega Thurber et al., 2008; Ng et al., 2009a,b]; representing dsDNA, ssDNA and RNA-containing viruses. While most viral metagenomic studies have been performed on purified virus particles, analyses of viral sequences present within microbial metagenomes have also been reported [Venter et al., 2004; DeLong et al., 2006; Williamson et al., 2008b].

Assembly-based estimations of dsDNA viral genotypic diversity vary substantially across marine ecosystems. For example, viruses collected from the Arctic Ocean are significantly less diverse than those collected from coastal waters off of British Columbia (∼500 genotypes vs. ∼130,000 genotypes) [Angly et al., 2006]. The upwelling regime that occurs off of the west coast of Canada was suggested as a possible explanation for the elevated levels of viral diversity in this area. Conversely, constraints on microbial diversity at high latitudes was likely responsible for depressed levels of viral diversity in the Arctic [Angly et al., 2006]. Viral communities extracted from marine sediment appear to be the most diverse, with up to an estimated 1 million genotypes per kilogram of sediment [Breitbart et al., 2004; Edwards and Rohwer, 2005]. This extremely high level of diversity may be in response to autochthonous microbial productivity in addition to allochthonous inputs from the overlying seawater. Metagenomic analysis of RNA-containing viruses inhabiting coastal marine ecosystems also revealed an unexpectedly diverse viral community, although the total number of genotypes was not estimated and therefore cannot be directly compared to other studies [Culley et al., 2006]. The RNA virome contained a diverse array of novel viral sequences that were only distantly related to known positive-sense ssRNA viral families.

Whole community sequencing of marine phage genomes using different approaches has resulted in conflicting theories regarding the biogeographical distribution of marine viruses. Metagenomic analysis of the viral particles collected from four oceanic regions suggested that marine viral “species” are globally distributed; yet the relative abundance of genotypes fluctuates between specific ecosystems [Angly et al., 2006]. These observations were based on how well viral sequences originating from different regions assembled with one another and subsequent estimations of richness, evenness, and abundance of genotypes. Despite the co-occurrence of phages in different oceanic regions, phylogenetic differences were also noted, suggesting geographical specificity. Alternatively, evaluation of viral sequences present within the microbial size fraction of metagenomic data collected during the Global Ocean Sampling (GOS) Expedition indicated that the spread of phage families, only myoviruses were ubiquitously distributed [Williamson et al., 2008b]. Assembled contigs that were attributed to podo- and siphoviruses were found to be more geographically isolated. While no significant correlations were found between the distribution of tailed phages and the environmental parameters that were measured, myoviruses appeared to be the most prevalent in tropical oligotrophic waters while podoviruses were more abundant in temperate coastal regions. It is likely that these conflicting theories in part stem from the different methods that were used to assess the occurrence and distribution of phages on a global basis.

Sequencing of individual viral genomes and viromes from the marine environment has unearthed a diverse range of virus-encoded cellular genes. The abundance and widespread global distribution of virus-encoded host genes was initially a shock to the microbial ecology research community. However, the high level of functional diversity witnessed within marine viromes is now becoming the norm (for a review of this topic, see Rohwer and Thurber [2009]). The complete genome sequences of several phages infective for two major cyanobacterial groups in the marine environment, Prochlorococcus and Synechococcus, offers perhaps the most striking example of how the presence of cellular genes within viral genomes can fundamentally alter our understanding of the importance of viruses to globally important biogeochemical processes [Mann et al., 2003, 2005; Sullivan et al., 2003, 2005, 2006; Lindell et al., 2004; Mann, 2005; Bryan et al., 2008]. The initial finding that cyanophages often carry photosystem I and II genes [Mann et al., 2003; Sullivan et al., 2006; Weigele et al., 2007; Sharon et al., 2009] has been followed by the discovery of a diverse array of cellular genes involved in metabolic and cellular processes ranging from phosphorus and carbon metabolism to nucleotide metabolism, vitamin B12 biosynthesis, antibiotic biosynthesis, virulence, and perhaps even regulation of programmed cell death [Mann, 2005; Sullivan et al., 2005; Weigele et al., 2007; Bryan et al., 2008]. The observed occurrence (and in some cases expression) [Lindell et al., 2005, 2007; Clokie
et al., 2006) of host genes suggests that these phages may influence the short-term adaptation of their hosts. In essence, these phages appear to be extending the lifespan of their hosts in an effort to increase replication efficiency. Metagenomic investigation of marine ecosystems has revealed an impressive abundance and distribution of cellular genes in viral communities [Sharont et al., 2007, 2009; Dinsdale et al., 2008; Williamson et al., 2008b]. Moreover, the metagenomic profiling of nine biomes, including the marine environment, [Dinsdale et al., 2008] suggests that microbial and even viral metagenomes are predictive of the biogeochemical conditions that characterize a particular environment. Together, these genomic and metagenomic investigations have revealed an intriguing first glimpse into the genetic details behind the impact of viral-encoded cellular genes on ecosystem processes of global importance.

2.3.2 Aquatic Environments: Freshwater

In contrast to the marine environment, only a few viral metagenomic studies have focused on freshwater ecosystems. Similar to marine studies, the viral community sequence data produced from a recreational lake in Maryland and a temporarily ice-covered lake in Antarctica are also quite unique [Dijkeng et al., 2009; Lopez-Bueno et al., 2009]. However, commonalities with marine viromes are more or less restricted to genetic novelty. Analysis of water samples collected from Lake Needwood in Maryland revealed the presence of predominantly RNA viral families that are known to infect a variety of higher organisms including plants, algae, insects, birds, and mammals with only one family specific to bacteria [Dijkeng et al., 2009]. Homologues to known pathogens of plants (i.e., Banana virus), insects (i.e., insect paralysis viruses), and mammals (i.e., circoviruses) were identified within assembled data, perhaps reflective of the various types of land usages and organisms that surround the lake. The Antarctic lake study conducted by Lopez-Bueno and co-workers produced significantly different results than most marine viromes, with an overrepresentation of eukaryotic viruses rather than phage [Lopez-Bueno et al., 2009]. The authors of this study witnessed a seasonal shift in viral community structure as the lake transitioned from an ice-covered to an open water system with eukaryotic ssDNA viruses dominating the former state and dsDNA phycoviruses (algal viruses) dominating the later. The abundance of ssDNA viral families that infect eukaryotes, such as mammals, birds and plants was unusual due to the absence of these organisms in and around the lake, suggesting that these viruses have evolved to infect different types of hosts. The emergence of phycoviruses that appeared in the lake following the summer thaw was most likely in response to a bloom of the prasinophytes, a green alga.

2.3.3 Terrestrial Environments

Fewer efforts have been directed toward investigating viral communities using metagenomic approaches in terrestrial environments. With respect to soil, this may be due in part to the difficulty in extracting a representative virus community [Williamson et al., 2005] and the daunting task of examining what is postulated to be the most diverse environment on Earth [Vogel et al., 2009]. For a more in-depth discussion of viruses in soils, see Chapter 4 in this volume. Metagenomic analysis of viruses that infect plants face similar challenges with respect to virus extraction, along with low virus titers [Lapidot et al., 2001]. However, metagenomic analyses of viruses from various terrestrial ecosystems have been produced, including desert, prairie, and rainforest soils [Fierer et al., 2007], rice paddy soil [Kim et al., 2008], and plants [Adams et al., 2009]. The first viral metagenomic investigation of soil, conducted by Fierer et al. [2007], was a relatively small study, with less than 5000 Sanger sequences analyzed across three soil types. In addition to targeted dsDNA virus community sequencing, the diversity of Bacteria, Archaea, and Fungi was also investigated through small-subunit rRNA amplification and sequencing. Estimations of virus diversity varied widely between soil types, ranging from ∼1000 genotypes in the desert sample to greater than 100 million genotypes in the rainforest sample. While the majority of viral sequences were unique and minimal overlap with marine and fecal viral metagenomes occurred, similarity to known phages of soil microbes was also observed.

In a different soil study targeting ssDNA viruses [Kim et al., 2008], a high level of genetic novelty was also observed with 60% of the sequences deemed unique. Known sequences were distantly related to several families of ssDNA viruses that infect eukaryotic organisms and were believed to originate from the plants resident to the sample plot (rice and others) as well as wild bird feces and composted manure. Viral metagenomic approaches have also been used as a diagnostic tool in plant virology [Adams et al., 2009]. For more details on this topic, see Chapter 7 in this volume. The authors of this study were able to successfully recover and detect seeded and novel RNA viral genomes from cDNA libraries that were dominated by a plant host signal using a combination of subtractive hybridization and pyrosequencing.

2.3.4 Extreme Environments

Viruses are abundant components of many “extreme” ecosystems such as hypersaline environments [Porter et al., 2007], deep-sea hydrothermal vents [Williamson