Contents

Preface xv

Contributors xvii

1. Introduction 1

Frans J. de Bruijn

Part 1 Background Chapters

2. DNA Reassociation Yields Broad-Scale Information on Metagenome Complexity and Microbial Diversity 5

Vigdis L. Torsvik and Lise Øvreås

3. Diversity of 23S rRNA Genes Within Individual Prokaryotic Genomes 17

Anna Pei, William E. Oberdorf, Carlos W. Nossa, Pooja Chokshi, Martin J. Blaser, Liying Yang, David M. Rosmarin, and Zhiheng Pei

4. Use of the rRNA Operon and Genomic Repetitive Sequences for the Identification of Bacteria 29

Andréa Maria Amaral Nascimento

5. Use of Different PCR Primer-Based Strategies for Characterization of Natural Microbial Communities 41

James I. Prosser, Shahid Mahmood, and Thomas E. Freitag

6. Horizontal Gene Transfer and Recombination Shape Mesorhizobial Populations in the Gene Center of the Host Plants *Astragalus Luteolus* and *Astragalus Ernestii* in Sichuan, China 49

Qiongfang Li, Xiaoping Zhang, Ling Zou, Qiang Chen, David P. Fewer, and Kristina Lindström

7. Amplified rDNA Restriction Analysis (ARDRA) for Identification and Phylogenetic Placement of 16S-rDNA Clones 59

Menachem Y. Sklarz, Roey Angel, Osnat Gillor, and Ines M. Soares

8. **Clustering-Based Peak Alignment Algorithm for Objective and Quantitative Analysis of DNA Fingerprinting Data** 67
Satoshi Ishii, Koji Kadota, and Keishi Senoo

Part 2 The Species Concept

9. **Population Genomics Informs Our Understanding of the Bacterial Species Concept** 77
Margaret A. Riley

10. **The Microbial Pangenome: Implications for Vaccine Development** 83
Annalisa Nuccitelli, Claudio Donati, Michèle A. Barocchi, and Rino Rappuoli

11. **Metagenomic Insights into Bacterial Species** 89
Konstantinos T. Konstantinidis

12. **Reports of Ad Hoc Committees for the Reevaluation of the Species Definition in Bacteriology** 99
Erko Stackebrandt

13. **Metagenomic Approaches for the Identification of Microbial Species** 105
David M. Ward, Melanie C. Melendrez, Eric D. Becraft, Christian G. Klatt, Jason M. Wood, and Frederick M. Cohan

Part 3 Metagenomics

14. **Microbial Ecology in the Age of Metagenomics** 113
Jianping Xu

15. **The Enduring Legacy of Small Subunit rRNA in Microbiology** 123
Susannah G. Tringe and Philip Hugenholtz

16. **Pitfalls of PCR-Based rRNA Gene Sequence Analysis: An Update on Some Parameters** 129
Erko Stackebrandt

17. **Empirical Testing of 16S PCR Primer Pairs Reveals Variance in Target Specificity and Efficacy not Suggested by *In Silico* Analysis** 135
Sergio E. Morales and William E. Holben

18. **The Impact of Next-Generation Sequencing Technologies on Metagenomics** 143
George M. Weinstock

19. **Accuracy and Quality of Massively Parallel DNA Pyrosequencing** 149
Susan M. Huse and David B. Mark Welch

20. **Environmental Shotgun Sequencing: Its Potential and Challenges for Studying the Hidden World of Microbes** 157
Jonathan A. Eisen

21. **A Comparison of Random Sequence Reads Versus 16S rDNA Sequences for Estimating the Biodiversity of a Metagenomic Sample** 163
Chaysavanh Manichanh, Charles E. Chapple, Lionel Frangeul, Karine Gloux, Roderic Guigo, and Joel Dore

22. **Metagenomic Libraries for Functional Screening** 171
Trine Aakvik, Rahmi Lale, Mark Liles, and Svein Valla

23. **GC Fractionation Allows Comparative Total Microbial Community Analysis, Enhances Diversity Assessment, and Facilitates Detection of Minority Populations of Bacteria** 183
William E. Holben

24. **Enriching Plant Microbiota for a Metagenomic Library Construction** 197
Ying Zeng, Hao-Xin Wang, Zhao-Liang Geng, and Yue-Mao Shen

25. **Towards Automated Phylogenomic Inference** 205
Martin Wu and Jonathan A. Eisen

26. **Integron First Gene Cassettes: A Target to Find Adaptive Genes in Metagenomes** 217
Lionel Huang and Christine Cagnon

27. **High-Resolution Metagenomics: Assessing Specific Functional Types in Complex Microbial Communities** 225
Ludmila Chistoserdova

28. **Gene-Targeted Metagenomics (GT Metagenomics) to Explore the Extensive Diversity of Genes of Interest in Microbial Communities** 235
Shoko Iwai, Benli Chai, Ederson da C. Jesus, C. Ryan Penton, Tae Kwon Lee, James R. Cole, and James M. Tiedje

29. **Phylogenetic Screening of Metagenomic Libraries Using Homing Endonuclease Restriction and Marker Insertion** 245
Torsten Thomas, Staffan Kjelleberg, and Pui Yi Yung

30. **ArrayOme- and tRNAcc-Facilitated Mobilome Discovery: Comparative Genomics Approaches for Identifying Rich Veins of Bacterial Novel DNA Sequences** 251
Hong-Yu Ou and Kumar Rajakumar

31. **Sequence-Based Characterization of Microbiomes by Serial Analysis of Ribosomal Sequence Tags (SARST)** 265
Zhongtang Yu and Mark Morrison

Part 4 Consortia and Databases

32. The Metagenomics of Plant Pathogen-Suppressive Soils 277
Jan Dirk van Elsas, Anna Maria Kielak, and Mariana Silvia Cretoiu

33. Soil Metagenomic Exploration of the Rare Biosphere 287
Tom O. Delmont, Laure Franqueville, Samuel Jacquiod, Pascal Simonet, and Timothy M. Vogel

34. The BIOSPAS Consortium: Soil Biology and Agricultural Production 299
Luis Gabriel Wall

35. The Human Microbiome Project 307
George M. Weinstock

36. The Ribosomal Database Project: Sequences and Software for High-Throughput rRNA Analysis 313
James R. Cole, Qiong Wang, Benli Chai, and James M. Tiedje

37. The Metagenomics RAST Server: A Public Resource for the Automatic Phylogenetic and Functional Analysis of Metagenomes 325
Elizabeth M. Glass and Folker Meyer

38. The EBI Metagenomics Archive, Integration and Analysis Resource 333
C. Hunter, G. Cochrane, R. Apweiler, S. Hunter

Part 5 Computer-Assisted Analysis

39. Comparative Metagenome Analysis Using MEGAN 343
Daniel H. Huson and Suparna Mitra

40. Phylogenetic Binning of Metagenome Sequence Samples 353
Alice Carolyn McHardy and Kaustubh Patil

41. Gene Prediction in Metagenomic Fragments with Orphelia: A Large-Scale Machine Learning Approach 359
Katharina H. Hoff, Maike Tech, Thomas Lingner, Rolf Daniel, Burkhard Morgenstern, and Peter Meinicke

42. Binning Metagenomic Sequences Using Seeded GSOM 369
Ching-Hung Tseng, Chon-Kit Kenneth Chan, Arthur L. Hsu, Saman K. Halgamuge, and Sen-Lin Tang

43. Iterative Read Mapping and Assembly Allows the Use of a More Distant Reference in Metagenome Assembly 379
Bas E. Dutilh, Martijn A. Huynen, Jolein Gloerich, and Marc Strous

44. Ribosomal RNA Identification in Metagenomic and Metatranscriptomic Datasets 387
Ying Huang, Weizhong Li, Patricia W. Finn, and David L. Perkins

45. SILVA: Comprehensive Databases for Quality Checked and Aligned Ribosomal RNA Sequence Data Compatible with ARB 393
Elmar Prüsse, Christian Quast, Pelin Yilmaz, Wolfgang Ludwig, Jörg Peplies, and Frank Oliver Glöckner

46. ARB: A Software Environment for Sequence Data 399
Ralf Westram, Kai Bader, Elmar Prüsse, Yadhu Kumar, Harald Meier, Frank Oliver Glöckner, and Wolfgang Ludwig

47. The Phyloware Project: A Software Framework for Phylogenomic Virtue 407
Daniel N. Frank and Charles E. Robertson

48. MetaSim: A Sequencing Simulator for Genomics and Metagenomics 417
Daniel C. Richter, Felix Ott, Alexander F. Auch, Ramona Schmid, and Daniel H. Huson

49. *ClustScan*: An Integrated Program Package for the Detection and Semiautomatic Annotation of Secondary Metabolite Clusters in Genomic and Metagenomic DNA Datasets 423
John Cullum, Antonio Starcevic, Janko Diminic, Jurica Zucko, Paul F. Long, and Daslav Hranueli

50. MetaGene: Prediction of Prokaryotic and Phage Genes in Metagenomic Sequences 433
Hideki Noguchi

51. Primers4clades: A Web Server to Design Lineage-Specific PCR Primers for Gene-Targeted Metagenomics 441
Bernardo Sachman-Ruiz, Bruno Contreras-Moreira, Enrique Zozaya, Cristina Martínez-Garza, and Pablo Vinuesa

52. A Parsimony Approach to Biological Pathway Reconstruction/Inference for Metagenomes 453
Yuzhen Ye and Thomas G. Doak

53. ESPRIT: Estimating Species Richness Using Large Collections of 16S rRNA Data 461
Yijun Sun, Yunpeng Cai, Li Liu, Fahong Yu, and William Farmerie

Part 6 Complementary Approaches

54. Metagenomic Approaches in Systems Biology 475
María-Eugenia Guazzaroni and Manuel Ferrer

55. Towards "Focused" Metagenomics: A Case Study Combining DNA Stable-Isotope Probing, Multiple Displacement Amplification, and Metagenomics 491
Yin Chen, Marc G. Dumont, Joshua D. Neufeld, and J. Colin Murrell

56. Suppressive Subtractive Hybridization Reveals Extensive Horizontal Transfer in the Rumen Metagenome 497
Elizabeth A. Galbraith, Dionysios A. Antonopoulos, Karen E. Nelson, and Bryan A. White

Part 6A Microarrays

57. GeoChip: A High-Throughput Metagenomics Technology for Dissecting Microbial Community Functional Structure 509
Joy D. van Nostrand, Zhili He, and Jizhong Zhou

58. Phylogenetic Microarrays (PhyloChips) For Analysis of Complex Microbial Communities 521
Eoin L. Brodie

59. Phenomics and Phenotype Microarrays: Applications Complementing Metagenomics 533
Barry R. Bochner

60. Microbial Persistence in Low-Biomass, Extreme Environments: The Great Unknown 541
Parag Vaishampayan, James N. Benardini, Myron T. La Duc, and Kasthuri Venkateswaran

61. Application of Phylogenetic Oligonucleotide Microarrays in Microbial Analysis 551
Pankaj Trivedi and Nian Wang

Part 6B Metatranscriptomics

62. Isolation of mRNA From Environmental Microbial Communities for Metatranscriptomic Analyses 569
Peer M. Schenk

63. Comparative Day/Night Metatranscriptomic Analysis of Microbial Communities in the North Pacific Subtropical Gyre 575
Rachel S. Poretsky and Mary Ann Moran

64. The "Double-RNA" Approach to Simultaneously Assess the Structure and Function of a Soil Microbial Community 587
Tim Urich and Christa Schleper

65. Soil Eukaryotic Diversity: A Metatranscriptomic Approach 597
Roland Marmeisse, Julie Bailly, Coralie Damon, Frédéric Lehembre, Marc Lemaire, Micheline Wésolowski-Louvel, and Laurence Fraissinet-Tachet

Part 6C Metaproteomics

66. Proteomics for the Analysis of Environmental Stress Responses in Prokaryotes 605
Ksenia J. Groh, Victor J. Nesatyy, and Marc J.-F. Suter

67. Microbial Community Proteomics 627
Paul Wilmes

68. Synchronicity between Population Structure and Proteome Profiles: A Metaproteomic Analysis of Chesapeake Bay Bacterial Communities 637
Jinjun Kan, Thomas E. Hanson, and Feng Chen

69. High-Throughput Cyanobacterial Proteomics: Systems-Level Proteome Identification and Quantitation 645
Saw Yen Ow and Phillip C. Wright

70. Protein Expression Profile of an Environmentally Important Bacterial Strain: The Chromate Response of *Arthrobacter* Species Strain FB24 663
Kristene L. Henne, Joshua E. Turse, Cindy H. Nakatsu, and Allan E. Konopka

Part 6D Metabolomics

71. The Small-Molecule Dimension: Mass-spectrometry-based Metabolomics, Enzyme Assays, and Imaging 677
Trent R. Northen

72. Metabolomics: High-Resolution Tools Offer to Follow Bacterial Growth on a Molecular Level 683
Lucio Marianna, Agnes Fekete, Moritz Frommberger, and Philippe Schmitt-Kopplin

73. Metabolic Profiling of Plant Tissues by Electrospray Mass Spectrometry 697
Heather Walker

74. Metabolite Identification, Pathways, and Omic Integration Using Online Databases and Tools 709
Matthew P. Davey

Part 6E Single-Cell Analysis

75. Application of Cytomics to Separate Natural Microbial Communities by their Physiological Properties 727

Susann Müller and David R. Johnson

76. Capturing Microbial Populations for Environmental Genomics 735

Martha Schattenhofer and Annelie Wendeberg

77. Microscopic Single-Cell Isolation and Multiple Displacement Amplification of Genomes from Uncultured Prokaryotes 741

Peter Westermann and Thomas Kvist

Index 747

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. Previously, the main problem 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 from 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 chronometers, 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 II: Metagenomics in Different Habitats*. Instead of relying only on a limited number of (long) review articles on selected topics, this book 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. Both 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 special feature of both books, facilitates the computer-assisted analysis of the vast amount of data generated by metagenomic studies. In addition, metagenomic studies in a variety of habitats are described, primarily in Volume II, which present a large number of system dependent different approaches in greatly differing habitats. The latter also results 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),

INRA the Institut National de Recherche de Agriculture (INRA), and

CNRS CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE the Centre National de Recherche Scientifique (CNRS).

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

FRANS J. DE BRUIJN

Castanet, Tolosan, France
March 2011