

Contents

Preface	ix
Acknowledgments	x
Abbreviations	xi
1 Introduction	1
1.1 Exploring the Microbial World	1
1.2 The DNA Sequencing Revolution: Historical Perspectives	4
References	7
2 The Architecture of Microbial Genomes	11
Introduction	11
2.1 Genome Size, Organization, and Replication	11
2.2 Nucleotide Composition	14
2.3 Ecological and Evolutionary Aspects of Microbial Genomes	16
2.3.1 The Role of Viruses in Promoting Genomic Diversity	18
2.4 Genomic Diversity in Microbial Communities	19
2.5 Does Genomic Diversity Matter?	21
References	21
3 Application of Omics Approaches to Earth and Environmental Sciences: Opportunities and Challenges	27
Introduction	27
3.1 New Perspectives on Microbial Biogeochemistry	27
3.1.1 Redefining the Carbon and Nitrogen Cycles	27
3.1.2 Omics as Sensitive and Efficient Tracers of Biogeochemical Processes	29
3.1.3 Omics Data is Valuable for Biogeochemical Models	29
3.1.4 Understanding Biotic Responses and Feedbacks to Global Change	29
3.2 A Genomic Record of Biological and Geochemical Evolution	30
3.3 Challenges and Limitations of Omics Approaches	32
3.4 Omics as a Complement to Other Approaches	33
References	34
4 Overview of Approaches: From Whole-Community Shotgun Sequencing to Single-Cell Genomics	41
Introduction	41
4.1 Choosing the Right Approach	41
4.1.1 Whole-Community Approaches	41

4.1.2	Targeted Approaches: Physical, Microbiological, and Isotopic Enrichment	43
4.1.3	Single-Cell Genomics	44
4.2	Experimental Design and Sampling Considerations	45
4.2.1	Replication	45
4.2.2	Estimating Sequencing Effort: How Much Sequencing Do I Need to Do?	46
4.2.3	From Sample to Data: Biases Due to Preservation, Storage, Extraction, and Sequencing	47
4.2.4	Estimating Absolute Abundance with Internal Standards	49
4.3	Overview of Current DNA Sequencing Technologies	49
4.4	Quality Control and Sequence Processing	51
4.4.1	Dereplication	51
4.4.2	Trimming	52
	References	53
5	Genomics of Single Species and Single Cells	59
	Introduction	59
5.1	Algorithms for Genome Assembly	60
5.2	Challenges of Genome Assembly	61
5.3	Scaffolding	63
5.4	Programs and Pipelines for Genome Assembly	63
5.5	Evaluation of Genome Assemblies	66
5.6	Single-Cell Genomics	67
	References	69
6	Metagenomics: Assembly and Database-Dependent Approaches	73
	Introduction	73
6.1	To Assemble or Not To Assemble?	73
6.2	Database-Dependent Approaches	75
6.3	Database-Independent Approaches: <i>De Novo</i> Assembly	78
6.4	Evaluation of Metagenomic Assemblies	82
6.5	A Philosophy of Metagenome Assemblies	82
	References	83
7	Metagenomic Binning	89
	Introduction	89
7.1	Genomic Signatures of Nucleotide Composition	90
7.2	Binning Programs	91
7.3	Additional Signal and Steps for Binning: Coverage, Taxonomic Data, and Mini-Assemblies	93
7.4	Identifying, Evaluating, and Assessing the Completeness of Genomic Bins	95
	References	97

8	Annotation: Gene Calling, Taxonomy, and Function	101
	Introduction	101
8.1	Gene Calling	102
8.2	Determining Taxonomic Composition	103
8.3	Functional Annotation	106
8.3.1	Overall Approach to Functional Annotation	106
8.3.2	Predicting Metabolic Pathways	107
8.3.3	The Importance of Experimental Annotation	108
	References	109
9	Metatranscriptomics	113
	Introduction	113
9.1	Sample Collection	114
9.2	RNA Extraction and Preparation of cDNA Libraries	115
9.2.1	Should rRNAs Be Removed Prior to Library Preparation and Sequencing?	115
9.3	Assigning Transcripts to Genes or Other Features	115
9.4	<i>De Novo</i> Assembly	116
9.5	Absolute Versus Relative Abundance and Normalization	118
9.6	Detecting Differential Expression	122
	References	123
10	Metaproteomics	127
	Introduction	127
10.1	Methodologies for Basic Proteomics	128
10.2	The Importance of Genomic Databases for Interpreting Proteomics Data	130
10.3	Quantitative Proteomics	131
10.4	Combining Stable Isotope Probing with Proteomics to Track Microbial Metabolism	133
	References	133
11	Lipidomics and Metabolomics	137
	Introduction	137
11.1	Lipidomics	137
11.2	Metabolomics	139
	References	140
12	Downstream and Integrative Approaches and Future Outlook	145
	Introduction	145
12.1	Comparative Omics	145
12.2	Statistical Approaches	146
12.3	Visualization	147
12.4	Cyberinfrastructure for Environmental Omics	148
12.4.1	Software Platforms for Integrated Analyses and Data Storage	149

12.5	Data and Sample Archival	151
12.6	Modeling	151
12.7	Emerging Trends and Future Outlook	153
	References	155
	Index	161