

# Contents

<b>Preface</b> .....	v
<b>Acknowledgements</b> .....	ix

## Part I Introductions

<b>1 Sequences as Biological Information: Cells Obey the Laws of Chemistry and Physics</b> .....	3
Why Study Microbes?.....	3
What is Biological Information and Where Does It Come From .....	5
How DNA Sequences Code for Information .....	7
From DNA to Protein: Transcription and Translation.....	9
DNA Sequences: More than Protein-Coding Genes .....	12
From DNA to DNA: Replication .....	14
Proteins: Structure and Function.....	14
<b>2 Bioinformatics for Microbiologists: An Introduction</b> .....	19
Identifying Similarities: Sequence Comparison by Means of Alignments .....	19
From Alignments to Phylogenic Relationships.....	28
Genome Annotation: the Challenge to Get It Right.....	31
Information Beyond the Single Genome .....	33
<b>3 Microbial Genome Sequences: A New Era in Microbiology</b> .....	37
The First Completely Sequenced Microbial Genome.....	37
The Importance of Visualization .....	38
Genome Atlases to Visualize Chromosomes .....	42
A Race Against the Clock: The Speed of Sequencing .....	44
The First Completely Sequenced Bacterial Genome .....	46
Comparative Bacterial Genomics .....	47
The Microbial Genome: Not All Bacteria Are Like <i>E. coli</i> .....	50
<b>4 An Overview of Genome Databases</b> .....	53
What is a Database? .....	54

Three Databases Storing Sequences and a Lot More..... 57

Data Files and Formats ..... 61

RNA Databases ..... 62

Protein Databases..... 64

**5 The Challenges of Programming: a Brief Introduction ..... 69**

Part 1: A Brief Overview of Computer Science Concepts..... 69

A Look at the Most Common Bioinformatic Procedures..... 73

Achieving Better Automation ..... 81

Part 2: Some Technical Details and Future Directions ..... 83

Programming Languages ..... 83

Markup Languages..... 86

Service Oriented Architecture..... 88

Specific Tools for Bioinformatic Use..... 89

**Part II Comparative Genomics**

**6 Methods to Compare Genomes: the First Examples ..... 95**

Genomic Comparisons: The Size of a Genome ..... 95

Pairwise Alignment of Genomes ..... 99

Comparing Gene Content and Annotation Quality ..... 100

RNA Comparisons: A Look at rRNAs ..... 102

Proteome Comparisons: What Makes a Family?..... 103

**7 Genomic Properties: Length, Base Composition and DNA Structures..... 111**

Length of Genomes: the ‘C-Value Paradox’ ..... 112

Genome Average Base Composition: The Percentage of AT ..... 114

GC Skew—Bias Towards The Replication Leading Strand ..... 118

Global Chromosomal Bias of AT Content ..... 122

DNA Structures ..... 125

The Structure Atlas..... 128

Bias In Purines—A-DNA Atlases..... 129

More on Structure Atlases..... 131

**8 Word Frequencies and Repeats ..... 137**

Analyzing Word Frequencies in a Genome..... 137

DNA Repeats Within a Chromosome ..... 139

Introduction to the DNA Repeat Atlas ..... 143

Local DNA Repeats are Related to Chromosomal AT Content ..... 146

DNA Structures Related to Repeats in Sequences..... 147

The Genome Atlas: Our Standard Method for Visualization ..... 147

**Part III Transcriptomics and Proteomics**

**9 Transcriptomics: Translated and Untranslated RNA**..... 153  
 Counting rRNA and tRNA Genes ..... 154  
 A Closer Look at Ribosomal RNA..... 155  
 Genes Encoding Transfer RNA..... 160  
 Genes Coding mRNA: Comparing Codon Usage Between Bacteria ..... 161  
 Other Non-coding RNA: tmRNA ..... 164

**10 Expression of Genes and Proteins** ..... 167  
 Comparing Gene Expression and Protein Expression ..... 168  
 Part 1: Regulation of Transcription..... 169  
 Part 2: Regulation of Translation ..... 179  
 Part 3: Protein Modification and Cellular Localization ..... 180  
 Antigen and Epitope Prediction ..... 185

**11 Of Proteins, Genomes, and Proteomes** ..... 189  
 Part 1: Analysis of Individual Protein-Coding Genes..... 190  
 Part 2: How to Annotate a Complete Genome ..... 197  
 Part 3: Proteome Comparisons..... 203

**PART IV MICROBIAL COMMUNITIES**

**12 Microbial Communities: Core and Pan-Genomics**..... 213  
 Defining Pan-Genomes and Core Genomes ..... 214  
 Current Data Available for Pan- and Core Genome Analysis..... 218  
 The Pan- and Core Genome of *Streptococcus* ..... 219  
 The Current *Bacillus* Pan- and Core Genome..... 221  
 An Overview of Some Proteobacterial Pan- and Core Genomes ..... 222  
 The *Burkholderia* Pan- and Core Genome..... 223

**13 Metagenomics of Microbial Communities**..... 229  
 Metagenomics Based on 16S rRNA Analysis..... 230  
 Metagenomics Based on Complete DNA Sequencing..... 232  
 Environmental Influences on Base Composition..... 234  
 Visualization of Environmental Metagenomic Data..... 235  
 Marine Metagenomics ..... 240  
 Other Metagenomics Applications..... 241

**14 Evolution of Microbial Communities; or, On the Origins of Bacterial Species** ..... 243  
 Where Does Diversity Come From?..... 244

Evolution Takes Time ..... 245

Evidence of Evolution in a Single Genome..... 247

Genome Islands..... 249

Evolution on a Chip ..... 252

Species and Speciation: *Vibrio cholerae*..... 253

Can We Predict Evolution? *Escherichia coli* Genome Reduction..... 253

  

**Abbreviations** ..... 257

  

**Index**..... 263