Contents

Numerical Examples xiii

1 Molecular Basis of Evolution 3
- 1.1. Evolutionary Tree of Life 3
- 1.2. Mechanism of Evolution 4
- 1.3. Structure and Function of Genes 5
- 1.4. Mutational Changes of DNA Sequences 9
- 1.5. Codon Usage 11

2 Evolutionary Change of Amino Acid Sequences 17
- 2.1. Amino Acid Differences and Proportion of Different Amino Acids 17
- 2.2. Poisson Correction (PC) and Gamma Distances 19
- 2.3. Bootstrap Variances and Covariances 25
- 2.4. Amino Acid Substitution Matrix 27
- 2.5. Mutation Rate and Substitution Rate 29

3 Evolutionary Change of DNA Sequences 33
- 3.1. Nucleotide Differences Between Sequences 33
- 3.2. Estimation of the Number of Nucleotide Substitutions 35
- 3.3. Gamma Distances 43
- 3.4. Numerical Estimation of Evolutionary Distances 45
- 3.5. Alignment of Nucleotide Sequences 46
- 3.6. Handling of Sequence Gaps in the Estimation of Evolutionary Distances 49

4 Synonymous and Nonsynonymous Nucleotide Substitutions 51
- 4.1. Evolutionary Pathway Methods 52
- 4.2. Methods Based on Kimura's 2-Parameter Model 62
- 4.3. Nucleotide Substitutions at Different Codon Positions 67
- 4.4. Likelihood Methods with Codon Substitution Models 69

5 Phylogenetic Trees 73
5.1. Types of Phylogenetic Trees 73
5.2. Topological Differences 81
5.3. Tree-Building Methods 83

6 Phylogenetic Inference: Distance Methods 87
6.1. UPGMA 87
6.2. Least Squares (LS) Methods 92
6.3. Minimum Evolution (ME) Method 99
6.4. Neighbor Joining (NJ) Method 103
6.5. Distance Measures to Be Used for Phylogenetic Reconstruction 111

7 Phylogenetic Inference: Maximum Parsimony Methods 115
7.1. Finding Maximum Parsimony (MP) Trees 116
7.2. Strategies of Searching for MP Trees 122
7.3. Consensus Trees 130
7.4. Estimation of Branch Lengths 131
7.5. Weighted Parsimony 133
7.6. MP Methods for Protein Data 138
7.7. Shared Derived Characters 140

8 Phylogenetic Inference: Maximum Likelihood Methods 147
8.1. Computational Procedure of ML Methods 147
8.2. Models of Nucleotide Substitution 152
8.3. Protein Likelihood Methods 159
8.4. Theoretical Foundation of ML Methods 162
8.5. Parameter Estimation for a Given Topology 163

9 Accuracies and Statistical Tests of Phylogenetic Trees 165
9.1. Optimization Principle and Topological Errors 165
9.2. Interior Branch Tests 168
9.3. Bootstrap Tests 171
9.4. Tests of Topological Differences 175
9.5. Advantages and Disadvantages of Different Tree-Building Methods 178

10 Molecular Clocks and Linearized Trees 187
10.1. Molecular Clock Hypothesis 187
10.2. Relative Rate Tests 191
10.3. Phylogenetic Tests 196
10.4. Linearized Trees 203

11 Ancestral Nucleotide and Amino Acid Sequences 207
11.1. Inference of Ancestral Sequences: Parsimony Approach 207
11.2. Inference of Ancestral Sequences: Bayesian Approach 208

11.3. Synonymous and Nonsynonymous Substitutions in Ancestral Branches 216
11.4. Convergent and Parallel Evolution 221

12 Genetic Polymorphism and Evolution 231
12.1. Evolutionary Significance of Genetic Polymorphism 231
12.2. Analysis of Allele Frequency Data 233
12.3. Genetic Variation in Subdivided Populations 236
12.4. Genetic Variation for Many Loci 244
12.5. DNA Polymorphism 250
12.6. Statistical Tests for Detecting Selection 258

13 Population Trees from Genetic Markers 265
13.1. Genetic Distance for Allele Frequency Data 265
13.2. Analysis of DNA Sequences by Restriction Enzymes 275
13.3. Analysis of RAPD Data 285

14 Perspectives 291
14.1. Statistical Methods 291
14.2. Genome Projects 292
14.3. Molecular Biology and Evolution 294

Appendices
A. Mathematical Symbols and Notations 297
B. Geological Timescale 299
C. Geological Events in the Cenozoic and Mesozoic Eras 301
D. Organismal Evolution Based on the Fossil Record 303

References 305

Index 329