

Contents

Preface *xi*
Acknowledgments *xiii*
Acronyms *xv*

1 What Is Biology? *1*
1.1 DNA: Nature, Role, and Function *3*
1.1.1 Nucleotides, Ligation, and Hybridization *4*
1.1.2 DNA: Its Role and Function *7*
1.1.2.1 Replication *8*
1.1.2.2 Transcription *9*
1.1.2.3 Translation *10*
1.1.3 Alignment-based Methods *11*
1.1.4 A Review of Common Alignment-free Methods *14*
1.1.4.1 Word-based Methods *16*
1.1.4.2 Information Content *17*
1.1.4.3 Entropy *18*
1.1.5 Major Results by Alignment-free Methods *18*
1.2 Hybridization Affinity *19*
1.2.1 Models of Gibbs Energy Landscapes *19*
1.2.2 Deep Structure of DNA *21*
1.2.3 New Alignment-free Methods from Deep Structure *23*
1.2.3.1 Noncrosshybridizing Bases *24*
1.2.3.2 Genomic Signatures *25*
1.2.3.3 Pmeric Signatures *28*
1.2.3.4 Genomic Information Systems *29*
1.3 Summary *30*
Problems *30*
Bibliography *31*

2	Genomes to Phenotypes, Environments, and Genomes	37
2.1	Phenotype Determination: The Conventional View	37
2.2	Environments and Autopoiesis	45
2.3	Phenotype Determination: An Alternative View	47
2.4	Summary	52
	Problems	53
	Bibliography	54
3	Species Definitions and Their Limitations	57
3.1	The Importance of the Concept of Species	57
3.2	Current Species Definitions	62
3.2.1	Morphological Species Concept	65
3.2.2	Biological Species Concept	66
3.2.3	Genetic Species Concept	67
3.2.4	Evolutionary Species Concept	67
3.2.5	Ecological Species Concept	67
3.2.6	Phylogenetic Species Concept	68
3.2.7	Recognition Species Concept	68
3.3	What Exactly Is a “Species”?	69
3.4	A Deeper Definition of “Species”	73
3.4.1	Biomarkers and Data Samples	77
3.4.2	A Geometric Atlas of Life	78
3.4.2.1	Molecular Delimitation of Species for Prokaryotes	79
3.4.2.2	Molecular Delimitation of Species for Eukaryotes	79
3.4.2.3	Simultaneous Molecular Delimitation of Species in Eukaryotes and Prokaryotes	79
3.4.2.4	Molecular Resolution of Species Complexes	81
3.4.3	Assessment of the Geometric Definition	81
3.5	Summary	84
	Problems	84
	Bibliography	85
4	Speciation and Evolution	89
4.1	Darwinian Evolution	89
4.1.1	Controversies on Darwin’s Perspective	93
4.2	Neo-Darwinian Evolution: The Modern Synthesis	94
4.3	Speciation Processes	97
4.3.1	Speciation Result of Geographic Barriers	99
4.3.1.1	Allopatric Speciation	99
4.3.1.2	Sympatric Speciation	100
4.3.1.3	Peripatric Speciation	101
4.3.1.4	Parapatric Speciation	102
4.3.2	Speciation from Genetic Divergences or Cytological Barriers	102
4.3.2.1	Sexual Selection	102
4.3.2.2	Genetic Drift	103

4.3.2.3	Polyploidy	104
4.3.2.4	Chromosomal Rearrangement	105
4.3.2.5	Recombinational Speciation	106
4.3.3	Controversial Issues in Speciation	107
4.4	Summary	108
	Problems	108
	Bibliography	109
5	From Genomes to Evolution	113
5.1	From Genotypes to Phenotypes: Evolutionary Scale	114
5.1.1	Examples of Microevolution	119
5.2	Is the Course of Evolution Predictable at the Macroscale?	120
5.2.1	Predictability in Evolution	121
5.2.2	How Does Evolution Work?	122
5.2.3	Limits on Predictability	124
5.3	Is the Course of Evolution Predictable at the Lineage Scale?	125
5.3.1	Case Study (Family): Cichlidae	125
5.3.1.1	Sample Data	126
5.3.1.2	Predictions on Evolutionary Trajectories	126
5.3.1.3	Discussion	126
5.3.2	Case Study (Genus): <i>Helicobacter</i>	129
5.3.2.1	Sample Data	130
5.3.2.2	Predictions on Evolutionary Trajectories	130
5.3.2.3	Discussion	131
5.3.2.4	Biological Assessment	132
5.4	Summary	134
	Problems	134
	Bibliography	135
6	Conventional Phylogenetics and Systematics	139
6.1	DNA: The Thin Thread of Life	140
6.1.1	Taxonomy, Phylogenetics, and Systematics	141
6.1.2	Evolutionary Trees	142
6.1.3	Molecular Clocks	143
6.1.4	The Effect of Horizontal Gene Transfer	145
6.1.5	Molecular Methods	145
6.2	Distance Methods versus Parsimony Analysis	146
6.2.1	Bootstrap and the Resampling Technique	147
6.3	Probabilistic Methods and Evolutionary Models	149
6.3.1	Evolutionary Substitution Models	150
6.3.2	Maximum Likelihood and Bayesian Inference	153
6.4	Case Study in Molecular Taxonomy: COI	155
6.4.1	Biology, Natural History, and Evolution of Black Flies	155
6.4.2	Molecular Taxonomy: The Case of the Black Flies	157
6.4.2.1	Data Samples	157

6.4.2.2	Intra- and Interspecific Divergence	160
6.4.2.3	Phylogenetic Inference	160
6.4.2.4	Maximum Parsimony	160
6.4.2.5	Maximum Likelihood Analysis	160
6.4.2.6	Bayesian Inference	160
6.4.2.7	Results	161
6.4.2.8	Relationships	161
6.4.2.9	Discussion	163
6.5	Summary	165
	Problems	166
	Bibliography	166
7	New Approaches to Phylogenetics	173
7.1	Phylogenetic Trees	173
7.1.1	Feature Frequency Profiles	175
7.1.2	Data Compression	175
7.1.3	Chaos Representation	176
7.1.4	Information Theory	176
7.2	Phylogenetic Tree Assessment	177
7.3	Phylogenetic Hypotheses Based on Deep DNA Structure	178
7.3.1	Discussion	180
7.4	Summary	180
	Problems	181
	Bibliography	181
8	About the Origin of Life	185
8.1	Theories of the Origin of Life on Earth: From Creationism to Panspermia	185
8.1.1	What Is Life?	186
8.1.2	Where/How Did Life Originate?	187
8.1.3	Theories of the Origin of Life	188
8.2	Complex Systems: A Holistic Approach	188
8.2.1	Dynamical Systems: Attractors and Basin Landscapes	189
8.2.2	DSs and Unpredictability	190
8.2.3	Self-assembly and Self-organization	194
8.3	A Case Study: Polytene Chromosomes in Black Flies	195
8.3.1	Chromosome Shape and Structure	196
8.3.1.1	What Are Polytene Chromosomes?	196
8.3.1.2	Architecture and Analysis of Polytene Chromosomes in Black Flies	196
8.3.1.3	Limitations of Polytene Chromosomes Biomarkers	200
8.3.2	Species Identification from Chromosome Structure	201
8.3.2.1	Data Gathering and Preprocessing	201
8.3.2.2	Image Segmentation into Bands	203
8.3.2.3	Digital Image Segmentation and Processing	205
8.3.3	An Application: Chromosome Decomposition	205
8.3.3.1	Feature Selection and Extraction	206
8.3.3.2	Training a Band Classifier as Solution	207

8.3.3.3	Quality of the Solution	208
8.4	Summary	210
	Problems	210
	Bibliography	211
9	Toward an Analytical Biology	217
9.1	What Is Data Science (DS)?	218
9.1.1	Classification Problems	218
9.1.2	Prediction Problems	220
9.1.3	Clustering Problems	222
9.2	What Is Machine Learning (ML)?	224
9.2.1	Decision Trees	224
9.2.2	Random Forests	224
9.2.3	<i>k</i> -nearest Neighbor	224
9.2.4	Neural Networks	225
9.2.5	Support Vector Machines	225
9.2.6	Self-organizing Maps	225
9.2.7	Voronoi Diagrams	225
9.2.8	<i>k</i> -means Clustering	225
9.3	Causality vs. Correlation	226
9.4	What Is Artificial Intelligence (AI)?	228
9.5	The Natural Intelligence of DNA and Biological Organisms	229
9.6	Summary	231
	Problems	231
	Bibliography	232
10	Appendices	235
10.1	Genotypes	235
10.2	Standard Biological Taxonomies	235
10.3	Divergence in Black Flies	238
10.4	Software	244
	Bibliography	244
	Index	245

