

Contents

Preface	xv
Using these notes	xv
Acknowledgments	xvi
List of algorithms	xix
 I Fundamentals	 1
1 History	2
1.1 Aristotle	2
1.2 Theophrastus	3
1.3 Pierre Belon	4
1.4 Carolus Linnaeus	4
1.5 Georges Louis Leclerc, Comte de Buffon	6
1.6 Jean-Baptiste Lamarck	7
1.7 Georges Cuvier	8
1.8 Étienne Geoffroy Saint-Hilaire	8
1.9 Johann Wolfgang von Goethe	8
1.10 Lorenz Oken	9
1.11 Richard Owen	9
1.12 Charles Darwin	9
1.13 Stammbäume	12
1.14 Evolutionary Taxonomy	14
1.15 Phenetics	15
1.16 Phylogenetic Systematics	16
1.16.1 Hennig's Three Questions	16
1.17 Molecules and Morphology	18
1.18 We are all Cladists	18
1.19 Exercises	19
 2 Fundamental Concepts	 20
2.1 Characters	20
2.1.1 Classes of Characters and Total Evidence	22
2.1.2 Ontogeny, Tokogeny, and Phylogeny	23
2.1.3 Characters and Character States	23
2.2 Taxa	26

2.3	Graphs, Trees, and Networks	28
2.3.1	Graphs and Trees	30
2.3.2	Enumeration	31
2.3.3	Networks	33
2.3.4	Mono-, Para-, and Polyphyly	33
2.3.5	Splits and Convexity	38
2.3.6	Apomorphy, Plesiomorphy, and Homoplasy	39
2.3.7	Gene Trees and Species Trees	41
2.4	Polarity and Rooting	43
2.4.1	Stratigraphy	43
2.4.2	Ontogeny	43
2.4.3	Outgroups	45
2.5	Optimality	49
2.6	Homology	49
2.7	Exercises	50
3	Species Concepts, Definitions, and Issues	53
3.1	Typological or Taxonomic Species Concept	54
3.2	Biological Species Concept	54
3.2.1	Criticisms of the BSC	55
3.3	Phylogenetic Species Concept(s)	56
3.3.1	Autapomorphic/Monophyletic Species Concept	56
3.3.2	Diagnostic/Phylogenetic Species Concept	58
3.4	Lineage Species Concepts	59
3.4.1	Hennigian Species	59
3.4.2	Evolutionary Species	60
3.4.3	Criticisms of Lineage-Based Species	61
3.5	Species as Individuals or Classes	62
3.6	Monoism and Pluralism	63
3.7	Pattern and Process	63
3.8	Species Nominalism	64
3.9	Do Species Concepts Matter?	65
3.10	Exercises	65
4	Hypothesis Testing and the Philosophy of Science	67
4.1	Forms of Scientific Reasoning	67
4.1.1	The Ancients	67
4.1.2	Ockham's Razor	68
4.1.3	Modes of Scientific Inference	69
4.1.4	Induction	69
4.1.5	Deduction	69
4.1.6	Abduction	70
4.1.7	Hypothetico-Deduction	71
4.2	Other Philosophical Issues	75
4.2.1	Minimization, Transformation, and Weighting	75
4.3	Quotidian Importance	76
4.4	Exercises	76

5	Computational Concepts	77
5.1	Problems, Algorithms, and Complexity	77
5.1.1	Computer Science Basics	77
5.1.2	Algorithms	79
5.1.3	Asymptotic Notation	79
5.1.4	Complexity	80
5.1.5	Non-Deterministic Complexity	82
5.1.6	Complexity Classes: P and NP	82
5.2	An Example: The Traveling Salesman Problem	84
5.3	Heuristic Solutions	85
5.4	Metricity, and Untrametricity	86
5.5	NP-Complete Problems in Systematics	87
5.6	Exercises	88
6	Statistical and Mathematical Basics	89
6.1	Theory of Statistics	89
6.1.1	Probability	89
6.1.2	Conditional Probability	91
6.1.3	Distributions	92
6.1.4	Statistical Inference	98
6.1.5	Prior and Posterior Distributions	99
6.1.6	Bayes Estimators	100
6.1.7	Maximum Likelihood Estimators	101
6.1.8	Properties of Estimators	101
6.2	Matrix Algebra, Differential Equations, and Markov Models	102
6.2.1	Basics	102
6.2.2	Gaussian Elimination	102
6.2.3	Differential Equations	104
6.2.4	Determining Eigenvalues	105
6.2.5	Markov Matrices	106
6.3	Exercises	107
II	Homology	109
7	Homology	110
7.1	Pre-Evolutionary Concepts	110
7.1.1	Aristotle	110
7.1.2	Pierre Belon	110
7.1.3	Étienne Geoffroy Saint-Hilaire	111
7.1.4	Richard Owen	112
7.2	Charles Darwin	113
7.3	E. Ray Lankester	114
7.4	Adolf Remane	114
7.5	Four Types of Homology	115
7.5.1	Classical View	115
7.5.2	Evolutionary Taxonomy	115

7.5.3	Phenetic Homology	116
7.5.4	Cladistic Homology	116
7.5.5	Types of Homology	117
7.6	Dynamic and Static Homology	118
7.7	Exercises	120
8	Sequence Alignment	121
8.1	Background	121
8.2	"Informal" Alignment	121
8.3	Sequences	121
8.3.1	Alphabets	122
8.3.2	Transformations	123
8.3.3	Distances	123
8.4	Pairwise String Matching	123
8.4.1	An Example	127
8.4.2	Reducing Complexity	129
8.4.3	Other Indel Weights	130
8.5	Multiple Sequence Alignment	131
8.5.1	The Tree Alignment Problem	133
8.5.2	Trees and Alignment	133
8.5.3	Exact Solutions	134
8.5.4	Polynomial Time Approximate Schemes	134
8.5.5	Heuristic Multiple Sequence Alignment	134
8.5.6	Implementations	135
8.5.7	Structural Alignment	139
8.6	Exercises	145
III	Optimality Criteria	147
9	Optimality Criteria—Distance	148
9.1	Why Distance?	148
9.1.1	Benefits	149
9.1.2	Drawbacks	149
9.2	Distance Functions	150
9.2.1	Metricity	150
9.3	Ultrametric Trees	150
9.4	Additive Trees	152
9.4.1	Farris Transform	153
9.4.2	Buneman Trees	154
9.5	General Distances	156
9.5.1	Phenetic Clustering	157
9.5.2	Percent Standard Deviation	160
9.5.3	Minimizing Length	163
9.6	Comparisons	170
9.7	Exercises	171

10 Optimality Criteria—Parsimony	173
10.1 Perfect Phylogeny	174
10.2 Static Homology Characters	174
10.2.1 Additive Characters	175
10.2.2 Non-Additive Characters	179
10.2.3 Matrix Characters	182
10.3 Missing Data	184
10.4 Edge Transformation Assignments	187
10.5 Collapsing Branches	188
10.6 Dynamic Homology	188
10.7 Dynamic and Static Homology	189
10.8 Sequences as Characters	190
10.9 The Tree Alignment Problem on Trees	191
10.9.1 Exact Solutions	191
10.9.2 Heuristic Solutions	191
10.9.3 Lifted Alignments, Fixed-States, and Search-Based Heuristics	193
10.9.4 Iterative Improvement	197
10.10 Performance of Heuristic Solutions	198
10.11 Parameter Sensitivity	198
10.11.1 Sensitivity Analysis	199
10.12 Implied Alignment	199
10.13 Rearrangement	204
10.13.1 Sequence Characters with Moves	204
10.13.2 Gene Order Rearrangement	205
10.13.3 Median Evaluation	207
10.13.4 Combination of Methods	207
10.14 Horizontal Gene Transfer, Hybridization, and Phylogenetic Networks	209
10.15 Exercises	210
11 Optimality Criteria—Likelihood	213
11.1 Motivation	213
11.1.1 Felsenstein's Example	213
11.2 Maximum Likelihood and Trees	216
11.2.1 Nuisance Parameters	216
11.3 Types of Likelihood	217
11.3.1 Flavors of Maximum Relative Likelihood	217
11.4 Static-Homology Characters	218
11.4.1 Models	218
11.4.2 Rate Variation	219
11.4.3 Calculating $p(D T, \theta)$	221
11.4.4 Links Between Likelihood and Parsimony	222
11.4.5 A Note on Missing Data	224
11.5 Dynamic-Homology Characters	224
11.5.1 Sequence Characters	225

11.5.2	Calculating ML Pairwise Alignment	227
11.5.3	ML Multiple Alignment	230
11.5.4	Maximum Likelihood Tree Alignment Problem	230
11.5.5	Genomic Rearrangement	232
11.5.6	Phylogenetic Networks	234
11.6	Hypothesis Testing	234
11.6.1	Likelihood Ratios	234
11.6.2	Parameters and Fit	236
11.7	Exercises	238
12	Optimality Criteria—Posterior Probability	240
12.1	Bayes in Systematics	240
12.2	Priors	241
12.2.1	Trees	241
12.2.2	Nuisance Parameters	242
12.3	Techniques	246
12.3.1	Markov Chain Monte Carlo	246
12.3.2	Metropolis–Hastings Algorithm	246
12.3.3	Single Component	248
12.3.4	Gibbs Sampler	249
12.3.5	Bayesian MC ³	249
12.3.6	Summary of Posterior	250
12.4	Topologies and Clades	252
12.5	Optimality versus Support	254
12.6	Dynamic Homology	254
12.6.1	Hidden Markov Models	255
12.6.2	An Example	256
12.6.3	Three Questions—Three Algorithms	258
12.6.4	HMM Alignment	262
12.6.5	Bayesian Tree Alignment	264
12.6.6	Implementations	264
12.7	Rearrangement	266
12.8	Criticisms of Bayesian Methods	267
12.9	Exercises	267
13	Comparison of Optimality Criteria	269
13.1	Distance and Character Methods	269
13.2	Epistemology	270
13.2.1	Ockham’s Razor and Popperian Argumentation	271
13.2.2	Parsimony and the Evolutionary Process	272
13.2.3	Induction and Statistical Estimation	272
13.2.4	Hypothesis Testing and Optimality Criteria	272
13.3	Statistical Behavior	273
13.3.1	Probability	273
13.3.2	Consistency	274
13.3.3	Efficiency	281
13.3.4	Robustness	282

13.4	Performance	282
13.4.1	Long-Branch Attraction	283
13.4.2	Congruence	285
13.5	Convergence	285
13.6	Can We Argue Optimality Criteria?	286
13.7	Exercises	287

IV Trees 289

14 Tree Searching 290

14.1	Exact Solutions	290
14.1.1	Explicit Enumeration	290
14.1.2	Implicit Enumeration—Branch-and-Bound	292
14.2	Heuristic Solutions	294
14.2.1	Local versus Global Optima	294
14.3	Trajectory Search	296
14.3.1	Wagner Algorithm	296
14.3.2	Branch-Swapping Refinement	298
14.3.3	Swapping as Distance	301
14.3.4	Depth-First versus Breadth-First Searching	302
14.4	Randomization	304
14.5	Perturbation	305
14.6	Sectorial Searches and Disc-Covering Methods	309
14.6.1	Sectorial Searches	309
14.6.2	Disc-Covering Methods	310
14.7	Simulated Annealing	312
14.8	Genetic Algorithm	316
14.9	Synthesis and Stopping	318
14.10	Empirical Examples	319
14.11	Exercises	323

15 Support 324

15.1	Resampling Measures	324
15.1.1	Bootstrap	325
15.1.2	Criticisms of the Bootstrap	326
15.1.3	Jackknife	328
15.1.4	Resampling and Dynamic Homology Characters	329
15.2	Optimality-Based Measures	329
15.2.1	Parsimony	330
15.2.2	Likelihood	332
15.2.3	Bayesian Posterior Probability	334
15.2.4	Strengths of Optimality-Based Support	335
15.3	Parameter-Based Measures	336
15.4	Comparison of Support Measures—Optimal and Average	336
15.5	Which to Choose?	339
15.6	Exercises	339

16 Consensus, Congruence, and Supertrees	341
16.1 Consensus Tree Methods	341
16.1.1 Motivations	341
16.1.2 Adams I and II	341
16.1.3 Gareth Nelson	344
16.1.4 Majority Rule	347
16.1.5 Strict	347
16.1.6 Semi-Strict/Combinable Components	348
16.1.7 Minimally Pruned	348
16.1.8 When to Use What?	350
16.2 Supertrees	350
16.2.1 Overview	350
16.2.2 The Impossibility of the Reasonable	350
16.2.3 Graph-Based Methods	353
16.2.4 Strict Consensus Supertree	355
16.2.5 MR-Based	355
16.2.6 Distance-Based Method	358
16.2.7 Supertrees or Supermatrices?	360
16.3 Exercises	361
 V Applications	 363
17 Clocks and Rates	364
17.1 The Molecular Clock	364
17.2 Dating	365
17.3 Testing Clocks	365
17.3.1 Langley-Fitch	365
17.3.2 Farris	366
17.3.3 Felsenstein	367
17.4 Relaxed Clock Models	368
17.4.1 Local Clocks	368
17.4.2 Rate Smoothing	368
17.4.3 Bayesian Clock	369
17.5 Implementations	369
17.5.1 r8s	369
17.5.2 MULTIDIVTIME	370
17.5.3 BEAST	370
17.6 Criticisms	370
17.7 Molecular Dates?	373
17.8 Exercises	373
 A Mathematical Notation	 374
Bibliography	376
Index	415

Color plate section between pp. 76 and 77