

Contents

Preface	xiii
Gilles DIDIER and Stéphane GUINDON	
Chapter 1. Trees: Combinatorics and Models	1
Gilles DIDIER and Stéphane GUINDON	
1.1. Introduction	1
1.2. Preliminary definitions	2
1.3. Counting trees	4
1.3.1. Fully labeled non-rooted trees	4
1.3.2. Binary trees with labeled leaves	6
1.3.3. Binary trees with labeled leaves and ordered internal nodes	7
1.3.4. Number of orders of internal nodes of a given tree	8
1.3.5. Directed binary trees	9
1.4. Probabilities of trees resulting from branching processes	9
1.5. Birth–death processes	12
1.5.1. Probability density of a birth–death tree	15
1.6. The coalescent	18
1.6.1. Links with “classical” models in population genetics	19
1.6.2. Moran’s model	19
1.6.3. The Wright–Fisher model	21
1.6.4. Generic model	21
1.6.5. Coalescent-generated tree probability density	23
1.7. Conclusion	23
1.8. References	25
Chapter 2. Models of Sequences and Discrete Traits Evolution	27
Étienne PARDOUX	
2.1. Introduction	27
2.2. Discrete set-valued continuous-time Markov process	28

2.2.1. Poisson processes	28
2.2.2. Finite set-valued continuous-time Markov process	29
2.3. Models of DNA sequence evolution	32
2.3.1. The Jukes–Cantor model	32
2.3.2. The Kimura model	33
2.3.3. The Felsenstein model	33
2.3.4. The HKY model	34
2.3.5. The general time reversible model	35
2.4. Models of rate evolution along the sequence	35
2.4.1. Independent and identically distributed rates along the sequence	35
2.4.2. Hidden Markov model	36
2.5. Models of discrete trait evolution	37
2.6. References	38
Chapter 3. Evolutionary Models of Continuous Traits	39
Paul BASTIDE, Mahendra MARIADASSOU and Stéphane ROBIN	
3.1. Motivations	39
3.1.1. Comparative methods	40
3.1.2. Studies of evolutionary phenomena	41
3.2. Brownian motion	42
3.2.1. Description	43
3.2.2. Phylogenetic regression and statistical transformations	44
3.2.3. Recursive algorithms for inference	47
3.3. Multivariate analysis	48
3.3.1. Description	48
3.3.2. Phylogenetic contrasts	49
3.3.3. Phylogenetic PCA	49
3.4. Gaussian models	51
3.4.1. Some limits of the Brownian motion	51
3.4.2. Ornstein–Uhlenbeck process	52
3.4.3. Biological interpretations and caveats	56
3.4.4. Further Gaussian processes	58
3.4.5. Heterogeneous evolution	60
3.4.6. Observation models	64
3.4.7. Model selection	66
3.5. Extensions and generalizations	67
3.5.1. Non-Gaussian models	67
3.5.2. Tree–trait interactions	67
3.5.3. Interactions between species	68
3.5.4. Trait of high dimension	69
3.6. Useful references	69

3.7. Acknowledgements	70
3.8. References	71
Chapter 4. Correlated Evolution: Models and Methods	79
Guillaume ACHAZ and Julien Y. DUTHEIL	
4.1. Introduction	79
4.2. Correlated evolution between traits	82
4.2.1. Species are not independent	82
4.2.2. The phylogenetically independent contrasts	84
4.2.3. Extending the linear model to account for phylogeny	86
4.2.4. Correlation between discrete traits	91
4.2.5. Examples of correlated traits	92
4.2.6. Jointly modeling traits and sequences	93
4.3. Correlated evolution within genomes	94
4.3.1. Within genes, between nucleotides	94
4.3.2. Within proteins, between amino acids	96
4.3.3. Within genomes, between genes	101
4.4. Genetics is also correlated evolution	103
4.4.1. In individuals	103
4.4.2. In pedigrees	104
4.4.3. In the population	106
4.5. Conclusion	109
4.6. References	110
Chapter 5. A Century of Genomic Rearrangements	117
Anne BERGERON and Krister M. SWENSON	
5.1. Introduction	117
5.2. Orderings of genes and the rearrangements that act on them	118
5.2.1. Basic representations and definitions	119
5.2.2. DCJ operations and the breakpoint graph	121
5.3. Counting DCJ scenarios	125
5.3.1. Scenarios for a balanced cycle of length $2m$	125
5.3.2. The (many) cycle decompositions of a breakpoint graph	126
5.4. Chromosomal contact data and weighted scenarios	129
5.4.1. A model incorporating chromosomal contacts	129
5.4.2. Planar trees and an algorithm for exploring them	131
5.4.3. Planar trees	132
5.5. Conclusion	136
5.6. References	138

Chapter 6. Phylogenetic Inference: Distance-Based Methods 141

Fabio PARDI

6.1. Introduction	141
6.2. Mathematical basis	143
6.3. Distance estimation	146
6.3.1. Estimating distances from aligned sequences	146
6.3.2. Other approaches to estimate distances	148
6.4. Tree inference	150
6.4.1. Fitting branch lengths with least squares	151
6.4.2. Scoring trees: from least squares to minimum evolution	153
6.4.3. NJ and other agglomerative algorithms	154
6.4.4. Beyond distances	157
6.5. Conclusion	158
6.6. References	159

Chapter 7. Computing Inference in Phylogenetic Trees 165

Laurent GUÉGUEN

7.1. Inferences and modeling	165
7.1.1. Inferences	165
7.1.2. Parsimony and likelihood	166
7.1.3. Maximum parsimony	166
7.2. Dynamic programming	169
7.2.1. Over the branches	170
7.2.2. Over the nodes	171
7.2.3. Over the tree	171
7.2.4. At the root	172
7.2.5. Recursion relations	172
7.2.6. Complexity reduction	174
7.2.7. Root management	175
7.3. Maximum parsimony	177
7.3.1. Ancestral interference	179
7.4. Likelihood	179
7.4.1. Root management	182
7.4.2. Computation at the nodes	182
7.4.3. Maximization, differentiation	184
7.4.4. Ancestral interference	188
7.5. References	190

Chapter 8. The Bayesian Paradigm in Molecular Phylogeny	193
Nicolas RODRIGUE	
8.1. Introduction	193
8.2. General principles of the Bayesian approach in phylogeny	194
8.2.1. Markov chain Monte Carlo sampling	197
8.2.2. Summary of posterior distribution and sampling	200
8.3. Demarginalization of the likelihood function	200
8.3.1. Parameter expansion	200
8.3.2. Data augmentation	202
8.4. Bayesian selection of substitution models	203
8.4.1. Relative model comparison <i>via</i> the Bayes factor	204
8.4.2. Absolute evaluation of models <i>via</i> predictive posterior simulation	206
8.5. Impacts and future directions	207
8.6. References	208
Chapter 9. Measures of Branch Support in Phylogenetics	213
Olivier GASCUEL and Frédéric LEMOINE	
9.1. Introduction	213
9.2. Local supports: parametric and non-parametric aLRT	215
9.2.1. Null branch test and its limitations	215
9.2.2. Local aLRT test, parametric version	217
9.2.3. Local aLRT test, SH-like nonparametric version	218
9.2.4. Comparison with an example of aLRT support and bootstrap	219
9.3. Phylogenetic bootstrap	221
9.3.1. Statistic bootstrap	221
9.3.2. The Felsenstein bootstrap	221
9.3.3. Transfer bootstrap	223
9.3.4. Comparison with an example of bootstrap supports	226
9.4. Bayesian supports	228
9.4.1. Principle, use of Markov Monte Carlo chains	228
9.4.2. Local Bayesian support	230
9.4.3. Comparison of Bayesian supports with an example	231
9.5. Discussion	231
9.6. References	234
Chapter 10. Fossils and Phylogeny	237
Michel LAURIN	
10.1. Inferences on topology	237
10.1.1. First approaches	237
10.1.2. Traits usable in paleontology	239

10.1.3. First quantitative approach: phenetics	241
10.1.4. Stratophenetics	242
10.1.5. Cladistics	242
10.1.6. Model-based approaches: likelihood, Bayesian approaches . . .	243
10.1.7. Fossils and molecular data	245
10.2. Dating the tree of life	245
10.2.1. First qualitative approaches	245
10.2.2. First statistical approaches	246
10.2.3. Molecular dating	247
10.2.4. Tip dating	249
10.2.5. Birth–death model-based dating	250
10.3. Conclusion	252
10.4. References	252

Chapter 11. Phylodynamics 259

Samuel ALIZON

11.1. Reconciling ecology, evolution and mathematics	259
11.2. Data and processors	260
11.2.1. New generation sequencing	261
11.2.2. PCR and capture	261
11.3. Infection phylogenies	262
11.3.1. Link to transmission chains	262
11.3.2. Dating and evolutionary rates	263
11.3.3. Biological applications of time calibration	264
11.4. Phylodynamics	265
11.4.1. A field in search of definition	265
11.4.2. As closely as possible to epidemiology	266
11.4.3. Coalescent	267
11.4.4. Birth–death models	269
11.4.5. Limitations of likelihood approaches	269
11.4.6. ABC phylodynamics	270
11.5. Infection phylogeography	271
11.6. Infection and viral life history traits	272
11.7. Perspectives and challenges	273
11.8. References	275

Chapter 12. Inference of Demographic Processes in Human Populations 283

Frédéric AUSTERLITZ

12.1. Introduction	283
12.2. Demographic inferences from population genetics data	286

12.2.1. Reconstruction of the history of Central African Pygmies	286
12.2.2. Inference of the history of populations in Central Asia	288
12.2.3. Impact of lifestyle on population growth dynamics	288
12.3. Inferring human expansions from next-generation sequence data	291
12.4. Reconstructing population dynamics from genetic and cultural data	294
12.4.1. Simultaneous analysis of genetic and linguistic diversity	294
12.4.2. Detecting the intergenerational transmission of reproductive success	295
12.5. Conclusion	296
12.6. References	297
List of Authors	303
Index	305