

CONTENTS

List of Contributors

xxvi

I Evolution in Populations	1
1 Trees of genes in populations	3
<i>Joseph Felsenstein</i>	
1.1 Introduction	3
1.2 Effects of evolutionary forces on coalescent trees	7
1.2.1 Population growth	7
1.2.2 Migration	8
1.2.3 Coalescents with recombination	9
1.2.4 Natural selection	11
1.3 Inference methods	12
1.3.1 Earlier inference methods	13
1.3.2 The basic equation	13
1.3.3 Rescaling times	14
1.3.4 How many coalescent trees?	15
1.3.5 Monte Carlo integration	15
1.3.6 Importance sampling	15
1.3.7 Independent sampling	16
1.3.8 Correlated sampling	18
1.3.9 Sampling from approximate distributions	20
1.3.10 Ascertainment and SNPs	20
1.3.11 Bayesian samplers	21
1.3.12 Future extensions	21
1.4 Programmes	23
1.5 The wave of the future	25
2 The evolutionary analysis of measurably evolving populations using serially sampled gene sequences	30
<i>Allen Rodrigo, Gregory Ewing, and Alexei Drummond</i>	
2.1 Introduction	30
2.2 Constructing phylogenetic trees from serially sampled data	33
2.2.1 Estimation of the expected number of substitutions in each interval, or a uniform substitution rate	34

2.2.2	Correction of pairwise distances	37
2.2.3	Clustering using UPGMA	37
2.2.4	Trimming back branches	37
2.2.5	sUPGMA and serial sample miscellany	38
2.3	Maximum-likelihood estimation of evolutionary rates	39
2.3.1	Single rate dated tips	39
2.3.2	Multiple rates dated tips	39
2.3.3	A few last words about likelihood and serial samples	42
2.4	The serial coalescent	44
2.5	Estimating population size and substitution rates under the s-coalescent	47
2.5.1	Changing population sizes and skyline plots	50
2.6	Estimating migration rates	52
2.7	Where to next?	54
II	Models of sequence evolution	63
3	Modelling the variability of evolutionary processes	65
	<i>Olivier Gascuel and Stephane Guindon</i>	
3.1	Introduction	65
3.1.1	Among-site heterogeneity	66
3.1.2	Mixing among-site and time-dependent variability	67
3.2	Mathematical tools and concepts	68
3.2.1	Markovian models of sequence evolution: the basis and assumptions	68
3.2.2	Neyman (two-state, DNA), GTR (DNA), WAG (protein), and NY ¹ (codon) models	72
3.2.3	Trees and likelihood calculations	75
3.2.4	Accounting for among-site variability using mixture models	76
3.2.5	Gamma-based rate across sites models and NY ³ (codon) models	78
3.2.6	Accounting for among-site and time variability using Markov-modulated Markov (MMM) models	79
3.2.7	On/Off (two-state, DNA), covarion-like (DNA) and compound codon models	82
3.3	Biological data sets	84
3.3.1	The role of Deficiens and Globosa genes in flower development	84
3.3.2	The singular dynamics of the envelope gene evolution during HIV-1 infection	85
3.4	The models in action: analysis of protein coding sequences	86
3.4.1	Among-site heterogeneity	87
3.4.2	Application: classification of sites into selection regimes	91

3.4.3	Among-site and lineage heterogeneity in a unified framework	94
3.4.4	Application: visualization of time-dependent variations at individual sites	96
3.5	Discussion	99
4	Phylogenetic invariants	108
	<i>Elizabeth S. Allman and John A. Rhodes</i>	
4.1	Introduction	108
4.2	Phylogenetic models on a tree	113
4.3	Edge invariants and matrix rank	115
4.4	Vertex invariants and tensor rank	118
4.5	Algebraic geometry and computational algebra	121
4.6	Invariants for specific models	126
4.6.1	Group-based models	126
4.6.2	The general Markov model	128
4.6.3	The strand symmetric model	129
4.6.4	Stable base distribution models	130
4.7	Invariants and statistical tests	131
4.8	Invariants and maximum-likelihood	132
4.9	Invariants and identifiability of complex models	135
4.10	Other directions	139
4.10.1	A tree construction algorithm	139
4.10.2	Invariants for gene order models	140
4.11	Concluding remarks	141
III	Tree shape, speciation, and extinction	147
5	Some models of phylogenetic tree shape	149
	<i>Arne Ø. Mooers, Luke J. Harmon, Michaël G. B. Blum, Dennis H. J. Wong, and Stephen B. Heard</i>	
5.1	Introduction	149
5.2	Background	150
5.3	Yule and Hey models	151
5.4	$\lambda = \text{function}(\text{trait})$	153
5.5	$\lambda = \text{function}(\text{age})$	154
5.6	$\lambda = \text{function}(\text{time})$	156
5.7	The neutral model	157
5.8	$\lambda = \text{function}(N)$	160
5.9	Concluding remarks	162
5.10	Appendix	163

6 Phylogenetic diversity: from combinatorics to ecology	171
<i>Klaas Hartmann and Mike Steel</i>	
6.1 Introduction and terminology	171
6.2 Definitions and combinatorial properties	172
6.2.1 The strong exchange property	174
6.2.2 Generalized Pauplin formula	174
6.2.3 Exclusive molecular phylodiversity	175
6.3 Biodiversity conservation	175
6.3.1 Simple indices	177
6.3.2 Noah's Ark Problem	178
6.3.3 Conservation time scale	181
6.3.4 Further algorithmic results	182
6.3.5 Extensions to the NAP	183
6.4 Loss of phylogenetic diversity under extinction models	184
6.4.1 Relationship between PD and time under an extinction process	186
6.5 Tree reconstruction using PD	188
6.5.1 Tree reconstruction from PD -values over an abelian group	189
6.6 Concluding comments	192
 IV Trees from subtrees and characters	 197
 7 Fragmentation of large data sets in phylogenetic analyses	 199
<i>Michael J. Sanderson, Cécile Ané, Oliver Eulenstein, David Fernández-Baca, Junhyong Kim, Michelle M. McMahon, and Raul Piaggio-Talice</i>	
7.1 Introduction	199
7.2 Basic definitions	203
7.3 Strategies for handling fragmentation of data sets	205
7.3.1 Strategy 1. Post-processing collections of trees	205
7.3.2 Strategy 2. Pre-processing by grove identification	206
7.3.3 Strategy 3. Pre-processing by clustering or optimization strategies	211
7.4 Conclusions	213
 8 Identifying and defining trees	 217
<i>Stefan Grünewald and Katharina T. Huber</i>	
8.1 Introduction	217
8.2 From biology to mathematics	218
8.2.1 Evolutionary trees and X -trees	218
8.2.2 Characters and (partial) partitions	219

8.2.3	Homoplasy and displaying	220
8.2.4	Question (Q) restated	221
8.3	Defining trees in terms of chordal graphs	222
8.3.1	Partition intersection graphs and restricted chordal completions	222
8.3.2	Minimal restricted chordal completions and distinguishing edges	225
8.4	Defining trees in terms of closure rules	226
8.4.1	Quartet closure rules	228
8.4.2	Split closure rules	230
8.4.3	The semi-dyadic closure and homoplasy-free evolution	232
8.5	Identifying trees in terms of chordal graphs	234
8.5.1	Restricted chordal completions revisited	235
8.5.2	Strongly distinguishing	236
8.6	Identifying trees in terms of quartets	238
8.6.1	The quartet graph	238
8.6.2	Small identifying quartet sets	240
8.7	Conclusion	241
V	From trees to networks	245
9	Split networks and reticulate networks	247
	<i>Daniel H. Huson</i>	
9.1	Introduction	247
9.2	Consensus networks and super networks	249
9.3	Split networks from sequences and distances	255
9.4	Hybridization and reticulate networks	260
9.5	Recombination networks	267
10	Hybridization networks	277
	<i>Charles Semple</i>	
10.1	Introduction	277
10.1.1	Preliminaries	279
10.2	Hybridization networks	280
10.3	A characterization of MINIMUM HYBRIDIZATION	282
10.3.1	Rooted subtree prune and regraft operation and agreement forests	283
10.3.2	Characterizations of MINIMUM HYBRIDIZATION and MINIMUM RSPR	285
10.3.3	Comparing $d_{\text{rSPR}}(\mathcal{T}, \mathcal{T}')$ and $h(\mathcal{T}, \mathcal{T}')$	288
10.3.4	Algorithms for constructing rSPR sequences and hybridization networks from agreement forests	290

10.4 Algorithmic applications of agreement forests	291
10.4.1 Reduction rules	292
10.4.2 A simple divide-and-conquer algorithm for MINIMUM HYBRIDIZATION	295
10.4.3 Galled-trees	299
10.5 Recombination networks	301
10.6 Hybridization networks in real time	304
10.6.1 Temporal representations	304
10.6.2 Time-ordered rooted subtree prune and regraft operations	306
10.7 Computational complexity	308
10.8 Concluding remarks	309
Index	315

Luke J. Harmon

Biodiversity Centre

University of British Columbia, Vancouver, Canada

<http://www.biology.ubc.ca/~harmon/>

Allen Wilson Centre for Molecular Biology and Evolution, University of Auckland, New Zealand, and