Contents

Preface *page* ix
Acknowledgements x
Summary of most significant capabilities of BEAST 2 xi

Part I Theory 1

1 Introduction 3

1.1 Molecular phylogenetics 4
1.2 Coalescent theory 6
1.3 Virus evolution and phylodynamics 8
1.4 Before and beyond trees 8
1.5 Probability and Bayesian inference 10

2 Evolutionary trees 21

2.1 Types of trees 21
2.2 Counting trees 24
2.3 The coalescent 27
2.4 Birth–death models 36
2.5 Trees within trees 40
2.6 Exercise 43

3 Substitution and site models 44

3.1 Continuous-time Markov process 45
3.2 DNA models 46
3.3 Codon models 51
3.4 Microsatellite models 52
3.5 Felsenstein's likelihood 52
3.6 Rate variation across sites 54
3.7 Felsenstein's pruning algorithm 55
3.8 Miscellanea 57

4 The molecular clock 58

4.1 Time-trees and evolutionary rates 58
4.2 The molecular clock 58
4.3 Relaxing the molecular clock 60
4.4 Calibrating the molecular clock 65

5 Structured trees and phylogeography 68

5.1 Statistical phylogeography 68
5.2 Multi-type trees 69
5.3 Mugration models 71
5.4 The structured coalescent 71
5.5 Structured birth–death models 73
5.6 Phylogeography in a spatial continuum 73
5.7 Phylodynamics with structured trees 74
5.8 Conclusion 76

Part II Practice 77

6 Bayesian evolutionary analysis by sampling trees 79

6.1 BEAUti 80
6.2 Running BEAST 86
6.3 Analysing the results 89
6.4 Marginal posterior estimates 90
6.5 Obtaining an estimate of the phylogenetic tree 91
6.6 Visualising the tree estimate 94
6.7 Comparing your results to the prior 94

7 Setting up and running a phylogenetic analysis 97

7.1 Preparing alignments 97
7.2 Choosing priors/model set-up 100
7.3 Miscellanea 112
7.4 Running BEAST 114

8 Estimating species trees from multilocus data 116

8.1 Darwin's finches 116
8.2 Bayesian multispecies coalescent model from sequence data 119
8.3 *BEAST 119
8.4 SNAPP 123

9 Advanced analysis 127

9.1 Sampling from the prior 127
9.2 Serially sampled data 128

9.3 Demographic reconstruction 129
9.4 Ancestral reconstruction and phylogeography 134
9.5 Bayesian model comparison 135
9.6 Simulation studies 138

10 Posterior analysis and post-processing 139

10.1 Trace log file interpretation 140
10.2 Model selection 145
10.3 Troubleshooting 148

11 Exploring phylogenetic tree space 154

11.1 Tree space 154
11.2 Methods of exploring tree space 156
11.3 Tree set analysis methods 157
11.4 Summary trees 159
11.5 DensiTree 163

Part III Programming 167

12 Getting started with BEAST 2 169

12.1 A quick tour of BEAST 2 170
12.2 BEAST core: BEAST-objects and inputs 173
12.3 MCMC library 174
12.4 The evolution library 180
12.5 Other bits and pieces 182
12.6 Exercise 183

13 BEAST XML 184

13.1 What is XML? 184
13.2 BEAST file format and the parser processing model 186
13.3 An annotated example 190
13.4 Exercise 193

14 Coding and design patterns 195

14.1 Basic patterns 195
14.2 Input patterns 198
14.3 `initAndValidate` patterns 200
14.4 `CalculationNode` patterns 201
14.5 Common extensions 203
14.6 Tips 204
14.7 Known ways to get into trouble 205
14.8 Exercise 206

15 Putting it all together 207

15.1 Introduction 207
15.2 What is a package? 208
15.3 BEAUti 209
15.4 Variable selection-based substitution model package example 214
15.5 Exercise 219

References 220
Index of authors 240
Index of subjects 244