Contents

1 Introduction 1
- 1.1 Strategic Considerations 1
- 1.2 Notations 4
- 1.3 Preparing the Computer 5
  - 1.3.1 Installations 5
  - 1.3.2 Configurations 7

2 First Steps in R for Phylogeneticists 9
- 2.1 The Command Line Interface 9
- 2.2 The Data Structures 11
  - 2.2.1 Vector 11
  - 2.2.2 Factor 14
  - 2.2.3 Matrix 15
  - 2.2.4 Data Frame 16
  - 2.2.5 List 17
- 2.3 The Help System 18
- 2.4 Creating Graphics 19
- 2.5 Saving and Restoring R Data 20
- 2.6 Using R Functions 20
- 2.7 Repeating Commands 21
  - 2.7.1 Loops 21
  - 2.7.2 *Apply*-Like Functions 22
- 2.8 Exercises 23

3 Phylogenetic Data in R 25
- 3.1 Phylogenetic Data as R Objects 25
  - 3.1.1 The Class `"phylo"` (ape) 26
  - 3.1.2 The Class `"phylog"` (ade4) 27
  - 3.1.3 The Class `"matching"` (ape) 27
  - 3.1.4 The Class `"treeshape"` (apTreeshape) 28
- 3.2 Reading Phylogenetic Data 28

3.2.1 Phylogenies . . . 28
3.2.2 Reading Internet Tree Databases . . . 30
3.2.3 Molecular Sequences . . . 30
3.3 Writing Data . . . 33
3.4 Manipulating Data . . . 35
3.4.1 Basic Tree Manipulation . . . 35
3.4.2 Rooted *Versus* Unrooted Trees . . . 36
3.4.3 Dichotomous *Versus* Multichotomous Trees . . . 37
3.4.4 Summarizing and Comparing Trees . . . 38
3.4.5 Converting Objects . . . 39
3.4.6 Manipulating DNA Data . . . 40
3.5 Generating Random Trees . . . 44
3.6 Case Studies . . . 46
3.6.1 *Sylvia* Warblers . . . 46
3.6.2 Phylogeny of the Felidae . . . 50
3.6.3 Snake Venom Proteome . . . 52
3.6.4 Mammalian Mitochondrial Genomes . . . 55
3.6.5 Butterfly DNA Barcodes . . . 62
3.7 Exercises . . . 64

4 Plotting Phylogenies . . . 65
4.1 Simple Tree Drawing . . . 65
4.1.1 Annotating Trees . . . 71
4.1.2 Showing Clades . . . 80
4.2 Combining Plots . . . 83
4.3 Large Phylogenies . . . 89
4.4 Perspectives . . . 92
4.5 Exercises . . . 94

5 Phylogeny Estimation . . . 95
5.1 Distance Methods . . . 96
5.1.1 Calculating Distances . . . 96
5.1.2 Simple Clustering and UPGMA . . . 99
5.1.3 Neighbor-Joining . . . 100
5.2 Maximum Likelihood Methods . . . 100
5.2.1 Substitution Models: A Primer . . . 101
5.2.2 Estimation with Molecular Sequences . . . 106
5.2.3 Finding the Maximum Likelihood Tree . . . 110
5.2.4 DNA Mining with PHYML . . . 111
5.3 Bootstrap Methods and Distances Between Trees . . . 112
5.3.1 Resampling Phylogenetic Data . . . 113
5.3.2 Bipartitions and Computing Bootstrap Values . . . 115
5.3.3 Distances Between Trees . . . 118
5.3.4 Consensus Trees . . . 118
5.4 Molecular Dating . . . 119

5.5 Case Studies . . . 121
5.5.1 *Sylvia* Warblers . . . 121
5.5.2 Phylogeny of the Felidae . . . 125
5.5.3 Butterfly DNA Barcodes . . . 129
5.6 Perspectives . . . 131
5.7 Exercises . . . 131

6 Analysis of Macroevolution with Phylogenies . . . 133
6.1 Phylogenetic Comparative Methods . . . 133
6.1.1 Phylogenetically Independent Contrasts . . . 135
6.1.2 Phylogenetic Autoregression . . . 138
6.1.3 Autocorrelative Models . . . 139
6.1.4 Multivariate Decomposition . . . 142
6.1.5 Generalized Least Squares . . . 144
6.1.6 Generalized Estimating Equations . . . 147
6.1.7 Mixed Models and Variance Partitioning . . . 149
6.1.8 The Ornstein–Uhlenbeck Model . . . 151
6.1.9 Perspectives . . . 153
6.2 Estimating Ancestral Characters . . . 154
6.2.1 Continuous Characters . . . 155
6.2.2 Discrete Characters . . . 156
6.3 Analysis of Diversification . . . 160
6.3.1 Graphical Methods . . . 161
6.3.2 Birth–Death Models . . . 163
6.3.3 Survival Models . . . 167
6.3.4 Goodness-of-Fit Tests . . . 169
6.3.5 Tree Shape and Indices of Diversification . . . 170
6.4 Perspectives . . . 172
6.5 Case Studies . . . 173
6.5.1 *Sylvia* Warblers . . . 173
6.5.2 Phylogeny of the Felidae . . . 176
6.6 Exercises . . . 180

7 Developing and Implementing Phylogenetic Methods in R . 183
7.1 Features of R . . . 183
7.1.1 Object-Orientation . . . 183
7.1.2 Variable Definition and Scope . . . 185
7.1.3 How R Works . . . 186
7.2 Writing Functions in R . . . 187
7.3 Interfacing R with Other Languages . . . 189
7.3.1 Simple Interfaces . . . 189
7.3.2 Complex Interfaces . . . 190
7.4 Writing R Packages . . . 192
7.4.1 A Minimalist Package . . . 192
7.4.2 The Documentation System . . . 193

7.5 Performance Issues and Strategies 193

References 199

Index 209