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
1.4 Other Readings 8

2 First Steps in R for Phylogeneticists 9
2.1 The Command Line Interface 10
2.2 The Help System 12
2.3 The Data Structures 13
2.3.1 Vector 13
2.3.2 Factor 16
2.3.3 Matrix 17
2.3.4 Data Frame 19
2.3.5 List 20
2.4 Creating Graphics 21
2.5 Saving and Restoring R Data 22
2.6 Using R Functions 23
2.7 Repeating Commands 23
2.7.1 Loops 24
2.7.2 *Apply*-Like Functions 25
2.8 Exercises 26

3 Phylogenetic Data in R 29
3.1 Phylogenetic Data as R Objects 29
3.1.1 Trees 30
3.1.2 Networks 32
3.1.3 Splits 33
3.1.4 Molecular Sequences 34

3.1.5 Allelic Data 36
3.1.6 Phenotypic Data 37
3.2 Reading Phylogenetic Data 37
3.2.1 Phylogenies 37
3.2.2 Molecular Sequences 38
3.2.3 Allelic Data 41
3.2.4 Reading Data Over the Internet 42
3.3 Writing Data 44
3.4 Manipulating Data 47
3.4.1 Basic Tree Manipulation 47
3.4.2 Rooted *Versus* Unrooted Trees 49
3.4.3 Graphical Tree Manipulation 51
3.4.4 Dichotomous *Versus* Multichotomous Trees 51
3.4.5 Summarizing and Comparing Trees 52
3.4.6 Manipulating Lists of Trees 55
3.4.7 Molecular Sequences 56
3.4.8 Allelic Data 60
3.5 Converting Objects 61
3.6 Managing Labels and Linking Data Sets 62
3.7 Sequence Alignment 64
3.8 Case Studies 66
3.8.1 *Sylvia* Warblers 66
3.8.2 Mammalian Mitochondrial Genomes 70
3.8.3 Butterfly DNA Barcodes 78
3.9 Exercises 79

4 Plotting Phylogenies 81
4.1 Simple Tree Drawing 81
4.1.1 Node, Tip and Edge Annotations 87
4.1.2 Axes and Scales 96
4.1.3 Manual and Interactive Annotation 96
4.1.4 Showing Clades 99
4.1.5 Plotting Phylogenetic Variables 102
4.2 Combining Plots 103
4.2.1 Tree–Variable Coplot 103
4.2.2 Cophylogenetic Plot 108
4.3 Large Phylogenies 110
4.4 Networks 114
4.5 Data Exploration with Animations 117
4.6 Exercises 121

5 Phylogeny Estimation 123
5.1 Distance Methods 124
5.1.1 Calculating Distances 125
5.1.2 Exploring and Assessing Distances 130
5.1.3 Simple Clustering, UPGMA, and WPGMA 133
5.1.4 Neighbor-Joining 134
5.1.5 Extensions of Neighbor-Joining: UNJ and BIONJ 135
5.1.6 Minimum Evolution 135
5.2 Maximum Likelihood Methods 138
5.2.1 Substitution Models: A Primer 139
5.2.2 Estimation with Molecular Sequences 146
5.2.3 Finding the Maximum Likelihood Tree 157
5.2.4 DNA Mining with PhyML and `modelTest` 159
5.3 Bayesian Methods 161
5.4 Other Methods 165
5.4.1 Parsimony 165
5.4.2 Hadamard Conjugation 167
5.4.3 Species Trees *Versus* Gene Trees 169
5.5 Bootstrap Methods and Distances Between Trees 170
5.5.1 Resampling Phylogenetic Data 171
5.5.2 Bipartitions and Computing Bootstrap Values 174
5.5.3 Distances Between Trees 178
5.5.4 Consensus Trees and Networks 179
5.6 Molecular Dating 182
5.6.1 Molecular Clock 183
5.6.2 Penalized Likelihood 184
5.6.3 Bayesian Dating Methods 185
5.7 Summary and Recommendations 187
5.8 Case Studies 189
5.8.1 *Sylvia* Warblers 190
5.8.2 Butterfly DNA Barcodes 199
5.9 Exercises 201

6 Analysis of Macroevolution with Phylogenies 203
6.1 Phylogenetic Comparative Methods 203
6.1.1 Phylogenetically Independent Contrasts 204
6.1.2 Phylogenetic Autocorrelation 209
6.1.3 Orthonormal Decomposition 213
6.1.4 Multivariate Methods 216
6.1.5 Generalized Least Squares 218
6.1.6 Generalized Estimating Equations 228
6.1.7 Mixed Models and Variance Partitioning 231
6.1.8 The Ornstein–Uhlenbeck Model 232
6.1.9 Phylogenetic Signal 236
6.1.10 Intraspecific Variation 239

6.1.11 Phylogenetic Uncertainty . . . 245
6.2 Estimating Ancestral Characters . . . 247
6.2.1 Continuous Characters . . . 248
6.2.2 Discrete Characters . . . 252
6.3 Analysis of Diversification . . . 258
6.3.1 Graphical Methods . . . 258
6.3.2 The Simple Birth–Death and Yule Models . . . 260
6.3.3 Time-Dependent Models . . . 264
6.3.4 Combining Phylogenetic and Taxonomic Data . . . 268
6.3.5 Trait-Dependent Models . . . 271
6.3.6 Other Methods . . . 277
6.3.7 Tree Shape and Indices of Diversification . . . 280
6.3.8 Tests of Diversification Shifts . . . 282
6.4 Ecology and Biogeography . . . 286
6.4.1 Phylogenetic Diversity . . . 286
6.4.2 Community Structure . . . 290
6.4.3 Phylogenetic Biogeography . . . 293
6.4.4 Niche and Bioclimatic Evolution . . . 297
6.4.5 Coevolutionary Phylogenetics . . . 299
6.5 Perspectives . . . 302
6.6 Case Studies . . . 302
6.6.1 *Sylvia* Warblers . . . 302
6.7 Exercises . . . 311

7 Simulating Phylogenies and Evolutionary Data . . . 313
7.1 Trees . . . 313
7.1.1 Simple Trees . . . 313
7.1.2 Coalescent Trees . . . 314
7.1.3 Speciation–Extinction Trees . . . 316
7.1.4 Tree Shapes . . . 319
7.2 Phenotypic Data . . . 320
7.2.1 Covariance-Based Simulation . . . 320
7.2.2 Time-Explicit Simulation . . . 324
7.3 Joint Tree–Trait Simulation . . . 327
7.4 Molecular Sequences . . . 328
7.5 Exercises . . . 330

8 Developing and Implementing Phylogenetic Methods in R . 331
8.1 Features of R . . . 331
8.1.1 Object-Orientation . . . 332
8.1.2 Variable Definition and Scope . . . 334
8.1.3 How R Works . . . 336
8.2 Writing Functions in R . . . 337
8.2.1 Programming Methods and Generics . . . 339
8.2.2 S3 *Versus* S4 . . . 341

8.3 Interfacing R with Other Languages 345
8.3.1 Simple Interfaces 345
8.3.2 Complex Interfaces 346
8.4 Writing R Packages 347
8.4.1 A Minimalist Package 348
8.4.2 The Documentation System 349
8.5 Performance Issues and Strategies 349
8.6 Computing with the class `"phylo"` 354
8.6.1 The Main Design and its Variants 354
8.6.2 Iterative *Versus* Recursive Computing 356

A Short Course on Regular Expressions 359

References 361

Index 381