Table of Contents

Chapter 1 ▪ Read Me First! 1

New and Improved Software 2
Just What Is a Phylogenetic Tree? 2
Estimating Phylogenetic Trees: The Basics 4
Beyond the Basics 5
Learn More about the Principles 7
About Appendix VI: F.A.Q. 7
Computer Programs and Where to Obtain Them 8
- *MEGA 7 8*
- *BEAST 8*
- *FigTree 8*
- *codeml 9*
- *SplitsTree and Dendroscope 9*
- *Graphviz 9*
- *Utility Programs 9*
- *Text Editors 9*

Acknowledging Computer Programs 9
The *Phylogenetic Trees Made Easy* Website 10

Chapter 2 ▪ Tutorial: Estimate a Tree 11

Why Create Phylogenetic Trees? 11
About this Tutorial 12
- *Macintosh and Linux users 12*
- *A word about screen shots 12*

Search for Sequences Related to Your Sequence 13
Decide Which Related Sequences to Include on Your Tree 18
  Establishing homology 19
  To include or not to include, that is the question 20
Download the Sequences 22
Align the Sequences 27
Make a Neighbor Joining Tree 28
Summary 32

Chapter 3 ▪ Acquiring the Sequences 33

Background 33
  Problems arising from the vast size of the sequence databases 33
  The query sequences 34
Hunting Homologs: What Sequences Can Be Included on a Single Tree? 34
Becoming More Familiar with BLAST 35
  BLAST help 36
Using the Nucleotide BLAST Page 36
Using BLAST to Search for Related Protein Sequences 39
Finalizing Selected Sequences for a Tree 43
  Problems adding coding sequences of protein homologs to the Alignment Explorer 43
Adding Sequences to and Removing Sequences from the Alignment Explorer 47
  Add a sequence 47
  Import a file of sequences 48
  Delete a sequence 48
Other Ways to Find Sequences of Interest (Beware! The Risks Are High) 48

Chapter 4 ▪ Aligning the Sequences 53

Aligning Sequences with MUSCLE 53
Examine and Possibly Manually Adjust the Alignment 57
  Trim excess sequence 57
  Eliminate duplicate sequences 58
Check Average Identity to Estimate Reliability of the Alignment 60
  Codons: Pairwise amino acid identity 60
  Non-coding DNA sequences 61
Increasing Alignment Speed by Adjusting MUSCLE's Parameter Settings 62
  How MUSCLE works 62
  Adjusting parameters to increase alignment speed 63

Aligning Sequences with ClustalW 64
Aligning Sequences with GUIDANCE2 65
  Viewing the results 67
  Eliminate unreliable parts of the alignment 68
  Saving the GUIDANCE2 alignment 70

Chapter 5 ▪ Major Methods for Estimating Phylogenetic Trees 71

LEARN MORE ABOUT TREE-SEARCHING METHODS 72
Distance versus Character-Based Methods 74
LEARN MORE ABOUT DISTANCE METHODS 74

Chapter 6 ▪ Neighbor Joining Trees 77

Using MEGA7 to Estimate a Neighbor Joining Tree 77
LEARN MORE ABOUT PHYLOGENETIC TREES 78
  Determine the suitability of the data for a Neighbor Joining tree 80
  Estimate the tree 81
LEARN MORE ABOUT EVOLUTIONARY MODELS 83
Unrooted and Rooted Trees 88
Estimating the Reliability of a Tree 91
LEARN MORE ABOUT ESTIMATING THE RELIABILITY OF PHYLOGENETIC TREES 92
What about Protein Sequences? 99

Chapter 7 ▪ Drawing Phylogenetic Trees 101

Changing the Appearance of a Tree 101
  The Options dialog 104
  Branch styles 108
  Fine-tuning the appearance of a tree 109
Rooting a Tree 112
  Finding an outgroup 112
Subtrees 114
Saving Trees 118
  Saving a tree description 118
  Saving a tree image 119
Captions 119

Chapter 8 ▪ Parsimony 121

LEARN MORE ABOUT PARSIMONY 121

MP Search Methods 123
Using SeaView for Parsimony 125
Estimating a bootstrap tree in SeaView 133
Using MEGA to draw the tree estimated by SeaView 136

Chapter 9 ▪ Maximum Likelihood 141
LEARN MORE ABOUT MAXIMUM LIKELIHOOD 141
ML Analysis Using MEGA 143
Test alternative models 144
Estimating the Reliability of an ML Tree by Bootstrapping 148
What about Protein Sequences? 149

Chapter 10 ▪ Bayesian Inference of Trees Using BEAST 151
BEAST: An Overview 151
Installing BEAST 152
Prepare the Input Alignment File 152
Run BEAUti 153
LEARN MORE ABOUT BAYESIAN INFERENCE 158
Running BEAST 162
Run Tracer 165
Burnin 166
Run TreeAnnotator 168
What about Protein Sequences? 169
Visualizing the BEAST Tree 170
The icons above the tree 172

Chapter 11 ▪ Which Method Should You Use? 175
Criteria to Consider 175
Accuracy 176
Ease of interpretation 176
Time and convenience 176
Results of the Major Methods 178

Chapter 12 ▪ Working with Various Computer Platforms 183
Command-line Programs 183
MEGA on the Macintosh Platform 184
Navigating among folders on the Mac 184
Printing trees and text from MEGA 188

The Line Endings Issue 188
Running the Utility Programs 189

Chapter 13 ▪ Phylogenetic Networks 191
Why Trees Are Not Always Sufficient 191
Unrooted and Rooted Phylogenetic Networks 192
LEARN MORE ABOUT PHYLOGENETIC NETWORKS 193
Using SplitsTree to Estimate Unrooted Phylogenetic Networks 198
  Estimating networks from alignments 198
  Rooting an unrooted network 204
  Estimating networks from trees 205
  Consensus networks 206
  Supernetworks 210
Using Dendroscope to Estimate Rooted Networks from Rooted Trees 211

Chapter 14 ▪ Minimum Spanning Trees 215
Minimum Spanning Trees Are *Not* Phylogenetic Trees! 215
Why Use Minimum Spanning Trees? 215
Origin of MSTs and the Issue of Reliability 216
  What is a minimum spanning tree? 216
Using MSTgold to Estimate MSTs 217
  The MSTgold input files 218
  Two ways for MSTgold to calculate the initial distance matrix 219
  Running MSTgold with the ebgC data 219
  The MSTgold output 220
Bootstrapping MSTgold 223
Exporting MSTs from Graphviz 225
An Alternative Data Set to Illustrate Some Additional Features of MSTgold 226
An Alternative to Graphviz: Hypercube 229

Chapter 15 ▪ Time Trees 231
Preparations to Estimate a Time Tree 231
Estimating a Time Tree 234
Viewing the Relative Time Tree 238
An Absolute Time Tree 240
  Effect of more calibration points on absolute time trees 244
Postscript 244

Chapter 16 ▪ Reconstructing Ancestral Sequences 245

Using MEGA to Estimate Ancestral Sequences by Maximum Likelihood 246
- *Create the alignment 246*
- *Construct the phylogeny 246*
- *Examine the ancestral states at each site in the alignment 247*
- *Estimate the ancestral sequence 250*
- *Calculating the ancestral protein sequence and amino acid probabilities 256*

How Accurate Are the Estimated Ancestral Sequences? 257

Chapter 17 ▪ Detecting Adaptive Evolution 259

Effect of Alignment Accuracy on Detecting Adaptive Evolution 261

Using MEGA to Detect Adaptive Evolution 261
- *Why* ebgC *is an interesting example for considering adaptive evolution 261*
- *Detecting overall selection 261*
- *Detecting selection between pairs 263*
- *Finding the region of the gene that has been subject to positive selection 264*

Using codeml to Detect Adaptive Evolution 266
- *Installation 267*
- *Run codeml 269*
- *Questions that underlie the models 273*
- *A closer look at ebgC_1.out 275*

Summary 278

Postscript 278
- *Compiling PAML yourself and running codeml without PAMLX 278*

Chapter 18 ▪ Estimating Phylogenetic Trees from Whole Genome Sequences 281

The Pan-Genome Problem 282

kSNP: An Alternative to Genome Alignment 282

kSNP 2.0 284
- *kChooser 284*
- *FCK 285*
- *Tree accuracy 286*
- *Always run kChooser before running kSNP 286*

Using kSNP3 286
- *The steps in estimating phylogenetic trees from WGS using kSNP3 287*

Chapter 19 ▪ Some Final Advice: Learn to Program 297

Appendix I ▪ File Formats and Their Interconversion 299

Format Descriptions 299
  The MEGA format 299
  The FASTA format 300
  The Nexus format 301
  The PHYLIP format 304
Interconverting Formats 305
  FastaConvert, MEGA, and SeaView 305

Appendix II ▪ Text Editors 307

Mac OS X: TextWrangler 307
Windows: Notepad++ 308
Linux: Gedit 308

Appendix III ▪ The Command-line Environment 311

Introduction and History 311
Terminal and Command Prompt: The Apps for Accessing the Command-line Environment 312
  The current directory 313
Entering Commands 313
Navigating in Terminal 314
  The Magic Trick 315
What Is in the CWD? The `ls` Command (§ `dir`) 316
  The -l option 317
  The -a option 317
  The permissions column 318
  The `chmod` *command 318*
§The `dir` Command in Command Prompt 319
Some Other Important Commands 320
  Copy a file 320
  Move a file 321
  Rename a file 321
  Make a directory 321
  Remove a directory 321
  Remove a file 321
  Print the contents of a file to the screen 321
  Clear the screen 321

Appendix IV ▪ Installing and Running Command-line Programs 323

Installing Command-line Programs 323
  Mac OS X / Linux 323
  Windows 324
Running Command-line Programs 329
  An example 330
  The line endings issue 330
  The example command line 330
  The output file(s) 330
  Error checking 330

Appendix V ▪ Additional Programs 331

Appendix VI ▪ Frequently Asked Questions 335

Literature Cited 339

Index to Major Program Discussions 341

Subject Index 345