

Table of Contents

Chapter 1 • Read Me First! 1

<i>New and Improved Software</i>	2
<i>Just What Is a Phylogenetic Tree?</i>	3
<i>Estimating Phylogenetic Trees: The Basics</i>	4
<i>Beyond the Basics</i>	5
<i>Learn More about the Principles</i>	6
<i>About Appendix III: F.A.Q.</i>	7
<i>Computer Programs and Where to Obtain Them</i>	7
MEGA	5 8
MrBayes	8
FigTree	8
Codeml	8
SplitsTree and Dendroscope	8
Utility Programs	8
Text Editors	9
<i>Acknowledging Computer Programs</i>	9
<i>The Phylogenetic Trees Made Easy Website</i>	9

Chapter 2 • Tutorial: Estimate a Tree 11

<i>Why Create Phylogenetic Trees?</i>	11
<i>About this Tutorial</i>	12
Macintosh and Linux users	12
A word about screen shots	12
<i>Search for Sequences Related to Your Sequence</i>	13
<i>Decide Which Related Sequences to Include on Your Tree</i>	16
Establishing homology	17
To include or not to include, that is the question	18

<i>Download the Sequences</i>	20
<i>Align the Sequences</i>	23
<i>Make a Neighbor Joining Tree</i>	24
<i>Summary</i>	28

Chapter 3 • Acquiring the Sequences 29

<i>Hunting Homologs: What Sequences Can Be Included on a Single Tree?</i>	29
<i>Becoming More Familiar with BLAST</i>	30
BLAST help	32
<i>Using the Nucleotide BLAST Page</i>	32
<i>Using BLAST to Search for Related Protein Sequences</i>	34
<i>Finalizing Selected Sequences for a Tree</i>	38
<i>Other Ways to Find Sequences of Interest (Beware! The Risks Are High)</i>	43

Chapter 4 • Aligning the Sequences 47

<i>Aligning Sequences with MUSCLE</i>	47
<i>Examine and Possibly Manually Adjust the Alignment</i>	51
Trim excess sequence	51
Eliminate duplicate sequences	54
<i>Check Average Identity to Estimate Reliability of the Alignment</i>	56
Codons: Pairwise amino acid identity	56
Non-coding DNA sequences	57
<i>Increasing Alignment Speed by Adjusting MUSCLE's Parameter Settings</i>	58
How MUSCLE works	58
Adjusting parameters to increase alignment speed	59
<i>Aligning Sequences with ClustalW</i>	60

Chapter 5 • Major Methods for Estimating Phylogenetic Trees 61

LEARN MORE ABOUT TREE-SEARCHING METHODS	62
<i>Distance versus Character-Based Methods</i>	64
LEARN MORE ABOUT DISTANCE METHODS	64
<i>Which Method Should You Use?</i>	66
Accuracy	66
Ease of interpretation	67
Time and convenience	67

Chapter 6 • Neighbor Joining Trees 69

Using MEGA 5 to Estimate a Neighbor Joining Tree 69

LEARN MORE ABOUT PHYLOGENETIC TREES 70

Determine the suitability of the data for a Neighbor Joining tree 73

Estimate the tree 74

LEARN MORE ABOUT EVOLUTIONARY MODELS 75

Unrooted and Rooted trees 80

Estimating the Reliability of a Tree 82

LEARN MORE ABOUT ESTIMATING THE RELIABILITY OF PHYLOGENETIC TREES 83

What about Protein Sequences? 89

Chapter 7 • Drawing Phylogenetic Trees 91

Changing the Appearance of a Tree 92

The Options dialog 94

Branch styles 96

Fine-tuning the appearance of a tree 99

Subtrees 102

Rooting a Tree 106

Finding an outgroup 108

Saving Trees 108

Saving a tree description 108

Saving a tree image 108

Captions 109

Chapter 8 • Parsimony 111

LEARN MORE ABOUT PARSIMONY 111

MP Search Methods 113

Multiple Equally Parsimonious Trees 116

Calculating branch lengths 117

Consensus and bootstrap trees 118

In the Final Analysis 122

Chapter 9 • Maximum Likelihood 123

LEARN MORE ABOUT MAXIMUM LIKELIHOOD 123

ML Analysis Using MEGA 125

Test alternative models 126

Rooting the ML tree 129

The special case of zero length branches	132
<i>Estimating the Reliability of an ML Tree by Bootstrapping</i>	134
<i>What about Protein Sequences?</i>	137

Chapter 10 • Bayesian Inference of Trees Using MrBayes 139

<i>MrBayes: An Overview</i>	139
LEARN MORE ABOUT BAYESIAN INFERENCE	141
Saving time (and perhaps your sanity)	142
Choose a model	143
<i>A General Strategy for Estimating Trees Using MrBayes</i>	143
<i>Creating the Execution File</i>	144
What the statements in the example mrbayes block do	145
How the stoprule option of the mcmc command is implemented	148
<i>How Do You Run a MrBayes Analysis?</i>	148
<i>More Complex (and More Useful) MrBayes Blocks</i>	149
Including a user tree	149
The nperts option of the mcmc command	150
Coding sequences and the charset statement	150
<i>The Screen Output while MrBayes Is Running</i>	151
<i>What If You Don't Get Convergence?</i>	152
<i>What about Protein Sequences?</i>	156
<i>Visualizing the MrBayes Tree</i>	156
<i>Using FigTree</i>	158
The side panel	158
The icons above the tree	160

Chapter 11 • Working with Various Computer Platforms 161

<i>Command Line Programs</i>	161
<i>MEGA on the Macintosh Platform</i>	162
Navigating among folders on the Mac	162
Printing trees and text from MEGA	165
<i>The Line Endings Issue</i>	165
<i>Installing Command Line Programs</i>	165
Macintosh and Linux: Use the bin folder	166
Windows: Create a bin folder and a path to it	166
<i>Command Line Programs: The Running Environment</i>	168

Windows: A brief visit to the Command Prompt program 168

Macintosh and Linux: A brief visit to Terminal and Unix 170

Acquiring and Installing MrBayes 172

Windows users 172

Macintosh and Linux users 173

Compile MrBayes for your Mac 173

Running the Utility Programs 174

Utility programs for Windows 175

Utility programs for Macintosh and Linux 175

Chapter 12 • Advanced Alignment Using GUIDANCE 177

Issues of Alignment Reliability 177

Unreliable sequences 177

Unreliable regions 178

How GUIDANCE Works 178

An Example Illustrated by the SmallData Data Set 179

Make a file of the unaligned sequences in FASTA format 180

Starting the run 180

Viewing the results 182

Eliminate unreliable sequences 186

Applications of GUIDANCE 190

Chapter 13 • Reconstructing Ancestral Sequences 191

Using MEGA to Estimate Ancestral Sequences by Maximum Likelihood 192

Create the alignment 192

Construct the phylogeny 193

Examine the ancestral states at each site in the alignment 194

Estimate the ancestral sequence 196

Calculating the ancestral protein sequence and amino acid probabilities 201

How Accurate are the Estimated Ancestral Sequences? 201

Chapter 14 • Detecting Adaptive Evolution 203

Effect of Alignment Accuracy on Detecting Adaptive Evolution 205

Using MEGA to Detect Adaptive Evolution 205

Detecting overall selection 205

Detecting selection between pairs 206

Finding the region of the gene that has been subject to positive selection 208

Using Codeml to Detect Adaptive Evolution 211

Installation 211

The files you need to run codeml	211
Questions that underlie the models	213
Run codeml	214
Identify the branches along which selection may have occurred	214
Test the statistical significance of the dN/dS ratios	216
<i>Summary</i>	218

Chapter 15 • Phylogenetic Networks 219

<i>Why Trees Are Not Always Sufficient</i>	219
<i>Unrooted and Rooted Phylogenetic Networks</i>	221
<i>Using SplitsTree to Estimate Unrooted Phylogenetic Networks</i>	221
Estimating networks from alignments	221
LEARN MORE ABOUT PHYLOGENETIC NETWORKS	223
Rooting an unrooted network	234
Estimating networks from trees	235
Consensus networks	236
Supernetworks	241
<i>Using Dendroscope to Estimate Rooted Networks from Rooted Trees</i>	243

Chapter 16 • Some Final Advice: Learn to Program 249

Appendix I • File Formats and Their Interconversion 251

<i>Format Descriptions</i>	251
The MEGA format	251
The FASTA format	252
The Nexus format	253
The PHYLIP format	256
<i>Interconverting Formats</i>	257
FastaConvert and MEGA	257
Other format conversion programs	257

Appendix II • Additional Programs 259

Appendix III • Frequently Asked Questions 263

Literature Cited	267
-------------------------	-----

Index to Major Program Discussions	269
---	-----

Subject Index	275
----------------------	-----