

Table of Contents

Chapter 1 Read Me First! 1

A Brief Overview 1

New and Improved Software 2

The Easy Steps to Making a Tree 2

Learn More about the Principles 4

Computer Programs Discussed and Where to Obtain Them 4

MEGA 5

PHYML 5

MrBayes 5

Codeml 5

Utility Programs 6

Macintosh Only! 6

The Phylogenetic Trees Made Easy Website 6

Register for Updated Information 7

Some Conventions Used in This Book 7

Chapter 2 Tutorial: Create a Tree! 9

Why Create Phylogenetic Trees? 9

About this Tutorial 10

MEGA 10

Obtaining Related Sequences by a BLAST Search 11

Step 1: Use BLAST to Search for Sequences Related
to Your Sequence 12

Step 2: Decide Which Related Sequences to Include on Your Tree 14

Step 3: Download the Sequence 18

Aligning the Sequences 22

Make a Neighbor-Joining Tree 25

Summary 29

Chapter 3 Acquiring and Aligning the Sequences 31

Selecting Homologs: What Sequences Can Be Put onto a Single Tree? 31

Finding Homologous Sequences by a BLAST Search 32

Other Ways to Find Sequences of Interest: Beware!

The Risks Are High 40

Aligning the Sequences 45

Eliminate Duplicate Sequences 51

Eliminate Obviously Wrong Sequences 53

Check Average Amino Acid Identity to Estimate Reliability of the Alignment 55

Non-Coding DNA Sequences 56

Interlude I Major Methods for Estimating Phylogenetic Trees 57

LEARN MORE ABOUT TREE-SEARCHING METHODS 58

Distance versus Character-Based Methods 60

Data Files Used to Illustrate Methods 61

Chapter 4 Neighbor-Joining Trees 63

Determine the Suitability of the Data for a Neighbor-Joining Tree 63

LEARN MORE ABOUT DISTANCE METHODS 65

LEARN MORE ABOUT PHYLOGENETIC TREES 67

Estimate the Neighbor-Joining Tree 70

LEARN MORE ABOUT EVOLUTIONARY MODELS 75

Estimating the Reliability of a Tree 80

LEARN MORE ABOUT ESTIMATING THE RELIABILITY
OF PHYLOGENETIC TREES 81

What about Protein Sequences? 88

Chapter 5 Drawing Phylogenetic Trees 89

Changing the Appearance of a Tree 89

Rooting a Tree 99

Finding an Outgroup 101

Subtrees 103

Saving Trees 105

Captions 106

Chapter 6 Parsimony 109

Search Methods 110

LEARN MORE ABOUT PARSIMONY 112

Multiple Equally Parsimonious Trees 115

Branch Lengths 117

Level of CNI Search 118

Consensus and Bootstrap Trees 118

What about Protein Sequences? 123

Chapter 7 Maximum Likelihood 125

Saving the Data in the PHYLIP Format 125

LEARN MORE ABOUT MAXIMUM LIKELIHOOD 126

Installing and Starting PHYML 129

Running PHYML 130

Importing the ML Tree Description into MEGA and Drawing the Tree 132

Tree Reliability Estimated by the approximate Likelihood Ratio Test Method (aLRT) 136

LEARN MORE ABOUT aLRT 137

The Special Case of Zero Length Branches 139

Test Alternative Models 140

What about Protein Sequences? 141

Chapter 8 Bayesian Inference of Trees with MrBayes 143

LEARN MORE ABOUT BAYESIAN ANALYSIS 144

Creating the Execution File 146

What the Statements in the Example MrBayes Block Do 148

What Happens during MrBayes Analysis? 149

The sumt Command 152

What If You Don't Get Convergence? 152

Visualizing the MrBayes Tree 154

What about Protein Sequences? 157

Forthcoming Version of MrBayes 157

Interlude II Which Method Should You Use? 159

Accuracy 160

Ease of Interpretation 160

Time and Convenience 161

Different Estimates of Tree Reliability 162

Chapter 9 Reconstructing Ancestral Sequences 165

Using MrBayes to Estimate Ancestral Sequences 166

The Alignment 166

Construct the Phylogeny 166

Estimate the Ancestral Sequence 166

Making Sense of the sump.anc File 169

Correcting Ancestral Sequences for Ancestral Gaps 173

The Emergency Sump Program 174

The Corrected Sequence 176

What about the Unreliable Sites? 176

Chapter 10 Detecting Adaptive Evolution 177

Using MEGA to Detect Adaptive Evolution 178

Using Codeml to Detect Adaptive Evolution 183

Windows Installation 183

Macintosh OSX Installation 183

Unix/Linux Users 183

The Files You Need to Run Codeml 183

The Codeml Control File 184

Questions That Underlie the Models 184

Run Codeml 186

Identify the Branches along Which Selection May Have Occurred 186

Test the Statistical Significance of the dN/dS Ratios 186

Chapter 11 Working with Various Computer Platforms 189

The Line Endings Issue 189

MEGA for Macintosh Users 190

PPC-Based Macs 190

Intel-Based Macs 191

Download and Install MEGA 191

MEGA for Unix/Linux Users 191

MEGA for Windows Users 191

Windows Users: Create a bin Folder and a Path to It 192

A Brief Visit to the Command Prompt Program for Windows Users 194

A Brief Visit to Unix for Macintosh Users 196

PHYML for Windows Users 197

PHYML for Macintosh Users 198

Install PHYML 198

PHYML for Unix/Linux Users 198

MrBayes for Windows Users 198

MrBayes for Macintosh Users 199

Compile MrBayes for Your Mac 199

Using the MrBayes Binary 199

<i>MrBayes for Unix/Linux Users</i>	200
<i>Utility Programs for Windows Users</i>	200
Installing the Utility Programs	200
Running the Utility Programs	200
<i>Utility Programs for Macintosh Users</i>	200
Installing the Utility Programs	200
Running the Utility Programs	201
<i>Utility Programs for Linux and Unix Users</i>	201
Compiling the Utility Programs	201
Running the Utility Programs	201
<i>Specific Instructions for the Utility Programs in the Download Packages</i>	202

Appendix I File Formats and Their Interconversion 205

<i>Formats Used by Programs Discussed in This Book</i>	205
The MEGA Format	205
The FASTA Format	206
The Nexus Format	207
PHYLIP	209
<i>Other File Formats</i>	211
The GCG/MSF Format	211
PileUp	211
The NBRF/PIR Format	212
<i>Interconverting Formats</i>	213

Appendix II Additional Programs 215

Literature Cited	221
-------------------------	------------

Index to Major Programs Discussed	223
------------------------------------------	------------

Subject Index	229
----------------------	------------