MacClade

Analysis of phylogeny
and character evolution

Version 3.08

Wayne P. Maddison
University of Arizona

David R. Maddison
University of Arizona

Sinauer Associates, Inc.
Publishers

SUNDERLAND, MASSACHUSETTS, U.S.A.

Copyright © 1992, 1999 by Wayne P. Maddison and David R. Maddison.
All rights reserved.

To order or for more information contact Sinauer Associates, Inc., Sunderland, Massachusetts 01375 U.S.A.

Names of products and companies mentioned in this book may be trademarks of their respective companies and are hereby acknowledged.

TABLE OF CONTENTS

Table of Contents

Preface

1. Introduction, 1

MacClade, 1
Learning how to use MacClade, 4
How to cite MacClade, 6
Earlier versions of MacClade, 6
Differences between MacClade 3 and 2.1, 7
Future plans for MacClade, 8
Electronic mail and Word Wide Web addresses, 9
Reporting bugs, 9
Technical information and availability of source code, 10

2. A Tutorial Overview of MacClade, 12

Tree manipulation, 12
Entering data, 18
Specifying assumptions, 20
Charting results, 20
Concluding remarks, 21

3. A Phylogenetic Perspective, 23

What is phylogeny?, 23
Comparisons broad and deep, 25
Speciation and extinction, 28
Character evolution: Adaptation, constraint, and rates, 30
Correlation between phylogenies, 33
Molecular evolution and population genetics, 34
Paleontology, 37
Outside of evolutionary biology, 37
Going both ways, 37

4. Introduction to Phylogenetic Inference, 39

Trees and their terms, 39
Characters and their terms, 43
Inferring the tree, 47
Character evolution: History versus models, 53
Inferring the history of character change, 54
Inferring models of character change using phylogenies, 54
What assumptions are acceptable?, 65
Parsimony versus other criteria, 69

5. Reconstructing Character Evolution Using Parsimony, 72
Assumptions about character evolution, 76
Finding the most-parsimonious ancestral states, 84
Finding equally parsimonious reconstructions, 105
Uncertainty in the reconstruction, 109
Statistics derived from reconstructions, 110
Locating and summarizing changes of character state, 112
Polytomies and their difficulties, 116
Uncertainty in the phylogenetic tree, 121

6. Stratigraphic Parsimony, 123
Introduction, 123
Concept of stratigraphic parsimony, 123
Definition of parsimony debt, 124
Nature of unit taxa, 124
Preliminary assessment and coding of stratigraphic data, 125
Computation of stratigraphic parsimony debt, 126
Stratocladistic analysis, 126
References, 128

7. Using MacClade on Your Macintosh, 130
Hardware and system software specifications, 130
MacClade and anti-viral software, 130
MacClade and the year 2000, 131
Installing MacClade, 131
Using MacClade with System 7 or later, or Multifinder, 131
Computer memory and MacClade's limits, 132
Starting MacClade, 133
Getting help, 134
Rules about selecting elements in a list, 134
Messages, 136
Setting preferences for MacClade's options, 136
Using MacClade with other programs, 138
Quitting from MacClade, 139

8. Managing Data Files, 140
Creating a new MacClade data file, 140
Opening an existing data file, 140
The NEXUS file format, 141
Adding notes about the data file, 142
Saving files to disk, 142
Saving files in other formats, 144
Altering the way MacClade writes files, 144
Reverting to last saved version, 144
Closing data files, 145

9. Importing and Exporting Text Files, 146
Importing files, 146
Exporting files, 148
MacClade 1.0 and 2.1 files, 149

PAUP files, 149
PHYLIP files, 150
HENNIG86 files, 152
NBRF files, 152
Simple table files, 153
Exporting descriptions, 155
Exporting MEGA files, 156

10. The Data Editor, 157
Creating and editing a data matrix, 158
Getting to the data editor, 160
Data formats, 160
Selecting and viewing elements in the editor, 162
Managing rows, columns, and blocks, 164
Names and symbols, 171
Entering data, 176
Proofreading your matrix, 185
Recoding character data, 185
Editor display options, 188
Annotating your data, 190

11. Assumptions About Characters, 194
Displaying and changing the current assumptions, 194
Character weighting and exclusion, 196
Protein-coding regions, 200
Character transformation assumptions, 200

12. Trees and Tree Manipulation, 210
Trees in MacClade, 210
MacClade's tree window, 211
Getting to the tree window, 211
Tree window structure and size, 212
Tree manipulation: Changing cladistic structure, 216
Tree display commands, 228
Tree storage and retrieval, 233
Adding notes about the trees, 235

13. Tracing Character Evolution, 236
Tracing the history of character evolution, 236
Interpreting the character tracing, 240
Display of character tracing, 244
Manipulating ancestral states, 246
Resolving equivocal tracings, 249
Character Data Boxes, 252
Summary of all changes on branches: Trace All Changes, 253
Branch lengths, 260
Summary of all states along each branch, 261
Confirming the validity of the parsimony algorithms, 261

14. Basic Tree and Character Statistics, 263

Display of statistics, 263
Number of steps, 264
Treelength, 265
Number of changes, 265
Tree changes, 266
Minimum and maximum possible steps, 267
Consistency and retention indices, 269

15. Charting Tree and Character Statistics, 272
General issues concerning chart calculations, 273
Character Steps/etc., 279
Treelengths of many trees, 285
State Changes & Stasis, 287
States, 296
Comparing two trees, 296
Comparing two tree files, 297
Chart display, 298
Confirming the validity of the charting calculations, 302

16. Continuous Characters, 303
To make discrete or not?, 303
Editing continuous characters, 304
Tracing continuous characters, 305
Confirming the validity of the parsimony algorithms, 306

17. Patterns of Correlated Character Evolution, 307
Using reconstructed histories to test correlation, 307
The concentrated-changes test, 308

18. Generating Random Data and Random Trees, 318
Random number generator and its seed, 318
Random data, 319
Random trees, 324

19. Using MacClade with Molecular Data, 327
DNA, RNA, and protein data formats, 327
Specifying coding regions and codon positions, 327
Translating nucleotide sequences to amino acid sequences, 329
Data editor features for sequence data, 330
Setting the genetic code, 334
Creating protein-parsimony user-defined types, 335
Importing data from molecular data bases, 336
File saving options, 336
Examples using molecular data, 336

20. Recording Your Work: Printing, Graphics, and Text Files, 343
Saving a record of your work, 343
Choosing fonts for printing, 345
Printing data and assumptions, 345
Printing trees, 346
Printing information for each branch, 355

Printing charts, 356
Saving graphics files, 356
Saving text files, 357

Appendix 1: File Format, 358
The NEXUS format, 358
Cursors, 359
Changing the way MacClade writes NEXUS files, 360

Appendix 2: Using MacClade and PAUP Together, 364
A scenario of using MacClade and PAUP, 364
Using both programs simultaneously, 365
Incompatibilities between MacClade and PAUP, 365
Differences in data file format, 372

References, 375

Annotated Index Organized by Menu and Palette, 386

Alphabetical Index, 396

PREFACE

MacClade began in 1985 when one of us (WPM) began to explore drawing and interacting with trees on the screen of an Apple Macintosh™ computer. From these initial explorations sprang MacClade version 1.0, which was released in June of 1986 (W. Maddison, 1986). It contained the central features of the Tree Window — namely, tracing of characters with fixing of the states at branches, rearrangement of branches of the tree using the computer's mouse, and a feature like Trace All Changes. It was primitive in other features.

MacClade version 2.1 was released in March of 1987 (Maddison and Maddison, 1987). It added the full Macintosh user interface of menus and dialog boxes, incorporated a built-in text editor, and added user-defined matrices of state-to-state distances ("step matrices").

Version 3 incorporates many new features beyond those found in version 2.1; some of the most noticeable are the specialized data editor, the chart window, acceptance of polytomous trees, and support for molecular data. For those of you who have used version 2.1, Chapter 1 gives a brief overview of some important differences between version 2.1 and version 3.

Following the original release of version 3.0 in 1992, several minor upgrades have been released. As of this writing, the current version is 3.08. The program is little changed since version 3.0, primarily differing in the addition of a few new features (for example, the ability to draw branch lengths proportional to reconstructed number of changes in Trace All Changes mode, an increase to 10 in the number of continuous characters allowed, and new random data generation options) and in bug fixes. This book differs from that published in 1992 only in the documentation of those few new features and in several corrections of typographical and similar errors.

Many people aided in the development of MacClade. We would like to thank H. W. Levi and E. O. Wilson, who gave us support and encouragement, and the freedom to work on MacClade when we should have been working on our theses. Some of the recent work on MacClade was done while WPM was a Natural Sciences and Engineering Research Council of Canada Post-doctoral Fellow, in Monty Slatkin's lab, and while DRM was a Alfred P. Sloan Foundation Post-doctoral Fellow, in R. C. Lewontin's lab; we thank them for their support.

We thank all those who reviewed pre-release copies of MacClade for us, and tested its functioning. Foremost among these was Willem Ellis, who searched long and hard for bugs in the program, and provided us with many thoughtful reports on MacClade. David Swofford, David Stern, Jeremy Ahouse, Darrel Frost, Ray Phillips, Robert O'Hara, Jenny Chappill, and Todd Disotell also deserve special note for their efforts in keeping us informed about MacClade bugs. Others who found bugs or gave suggestions include Belinda Chang, Jonathan Coddington, Chuck Crumly, Mike Cummings, Michael Donoghue, Jim Doyle, Sharon Gowan, Mike Hammer, Toby Kellogg, Jean-François Landry, Jim Malusa, Santiago Madriñan, Ted Oakes, Norm Platnick, and Jim Woolley. We are grateful for the encouragement we received from those testing the program, especially Michael Donoghue, David Swofford, Willem Ellis, and Mike Cummings. We thank the many users of MacClade 3.0–3.07 who sent us reports of difficulties and helped us to track down problems.

We would like to thank specifically several people who had an impact on the content of MacClade: David Swofford, who was always there for lively discussions about phylogenetics programs, and who provided programming advice, suggestions for improvement, and code for the random number generator; J. S. Ashe, whose interest in the charts summarizing statistics over multiple trees convinced us that it was worth the effort to include; Scott Edwards, whose desire to see the distribution of transitions and transversion over the length of a DNA sequence inspired the inclusion of Restrict Changes in the Character Steps charts; Allan Wilson and Linda Vigilant for inventing circular trees.

The book was improved by the inclusion of Dan Fisher's chapter on stratocladistics; we especially thank him for being willing to publish an outline of his ideas here. For comments, suggestions and discussion that contributed to this book, we thank Helen Amerongen, Leticia Avilés, David Baum, Michael Donoghue, Willem Ellis, Joe Felsenstein, Daniel Fisher, Larry Gall, Junhyong Kim, Michael Sanderson, and David Swofford.

We are very grateful to Michael Donoghue, Walter Fitch, Arnold Kluge, James Lake, Diana Lipscomb, and Mark Stoneking for supplying electronic versions of data files which form the basis of some of the example data files distributed with MacClade.

MacClade would not have arisen if it were not for Apple Computer's development of the computer and graphical interface that made it possible. We also thank them for supporting our work with a generous award.

Throughout the long gestation period of this current version of MacClade, Andy Sinauer has been incredibly patient and tolerant of

the many delays. We also thank Joe Vesely, Carol Wigg, Roberta Lewis, Kathaleen Emerson, and Dean Scudder for their efforts.

Finally, we would like to thank Leticia Avilés and Helen Amerongen, for their support through many long hours of programming, and for their endurance of horrendous phone bills.

Wayne P. Maddison
Department of Ecology
and Evolutionary Biology
University of Arizona
Tucson, AZ 85721

David R. Maddison
Department of Entomology
University of Arizona
Tucson, AZ 85721

Part I

Introducing MacClade