MacClade

for the Apple Macintosh ™

MacClade and the MacClade manual are Copyright 1986 by Wayne Maddison, all rights reserved except as follows. You may copy the program and documentation, as long you do not sell the copies. You may not modify the program or documentation.

You can become a registered user by purchasing the program directly from me. Please write me for information. Although prices are subject to change, I plan to keep the cost of the program at cost (less then $10). Registered users will receive notification of new versions and any substantial, non-cosmetic bugs.

Some necessary stuff:

LIMITED WARRANTY. Even though I (Wayne Maddison) have tested the software and reviewed the documentation, I make no warranty, either express or implied, with respect to this software, its quality, performance, merchantability, or fitness for a particular purpose. As a result, this software is sold "as is", and you, the user, are assuming the entire risk as to its quality and performance. Should the programs prove defective, the user assumes the entire risk, including any and all necessary servicing, repair or correction, as well as any and all incidental or consequential damages.

System, Finder, and Imagewriter Driver are copyrighted programs of Apple Computer, Inc. licensed to Wayne Maddison to distribute for use only in combination with MacClade. Apple Software shall not be copied onto another diskette (except for archive purposes) or into memory unless as part of the execution of MacClade. When MacClade has completed execution Apple Software shall not be used by any other program.

APPLE COMPUTER, INC. MAKES NO WARRANTIES, EITHER EXPRESS OR IMPLIED, REGARDING THE ENCLOSED COMPUTER SOFTWARE PACKAGE, ITS MERCHANTABILITY OR ITS FITNESS FOR ANY PARTICULAR PURPOSE. THE EXCLUSION OF IMPLIED WARRANTIES IS NOT PERMITTED BY SOME STATES. THE ABOVE EXCLUSION MAY NOT APPLY TO YOU. THIS WARRANTY PROVIDES YOU WITH SPECIFIC LEGAL RIGHTS. THERE MAY BE OTHER RIGHTS THAT YOU MAY HAVE WHICH VARY FROM STATE TO STATE.

Macintosh is a trademark licensed to Apple Computer, Inc. Apple is a registered trademark and MacWrite, MacPaint and Macintosh Pascal are trademarks of Apple Computer, Inc. Microsoft, Multiplan, and Excel are trademarks of Microsoft Corporation. MockWrite is a program distributed by CE Software.

MacClade

for the Apple Macintosh

Manual for version 1.0, June 1986

Wayne Maddison
Museum of Comparative Zoology, Harvard University, Cambridge MA 02138
(617) 495-2472

INTRODUCTION

MacClade was written to allow systematists and other evolutionary biologists to analyze character evolution in an interactive environment. Emphasis was placed on graphic display and ease of use. MacClade is therefore also a good teaching device for courses introducing phylogenetic principles.

MacClade has a number of features for a flexible and interactive analysis of character evolution. MacClade displays an input cladogram and reconstructs the evolution of characters. The evolution of a character is displayed by "painting" the cladogram with different shades (Figure 2). These shades indicate all of the most parsimonious reconstructions of ancestral states, according to the cladogram, but the user can examine other hypotheses of character evolution as well by fixing the state at a branch selected to a state of the user's choice. Besides allowing the user to explore alternative hypotheses under the same cladogram, MacClade allows quick comparisons of character evolution according to different cladograms, because the user can easily pick up and move branches on-screen. These features illustrate one of the motivations behind MacClade: the conviction that if evolutionary biologists are to test their hypotheses of evolutionary processes against phylogenetic patterns, the techniques for analyzing character evolution must be improved and made easier. The Pell. chromosomes data set, included on the MacClade disk, is being used in a study on spider chromsome evolution in which I explore questions such as how many times the XXXY sex chromosomes evolved, whether their evolution is phylogenetically correlated with the evolution of chiasma localization, how much more parsimonious is it to suppose independent gain of XXXY over gain then independent loss, and so on.

The ability to pick up and move branches on screen is one of several features that can be used to search for parsimonious cladograms. Clades can be rerooted at branches selected on the screen. Character weights and assumptions like irreversibility can be changed easily. Terminal taxa can be moved down into ancestral position, a feature useful to paleontologists wishing to examine the consequences of supposing a fossil is the actual ancestor. With each change, the new cladogram length and character evolution are shown. A user familiar with the organisms studied stands a good chance of finding parsimonious cladograms. However, this is not guaranteed, and MacClade is limited in its ability to find cladograms for the user: its only automatic cladogram-finding algorithm is local branch swapping. Nonetheless MacClade's interactiveness encourages the user's exploration of the data set, and insights can be gained that would contribute to an intelligent reconstruction of phylogeny. MacClade would be most profitably used

in combination with the more sophisticated cladogram-finding programs like PAUP, PHYLIP, and PHYSYS. MacClade could be used initially to test favorite cladograms, explore the data set, explore different assumptions, and then the data set could be sent through one of the above programs for a more powerful cladogram search. The user could then return to MacClade to explore the newly found cladograms. There are plans to put PAUP and possibly PHYLIP onto the Macintosh.

Specifications: MacClade is written in Pascal, with the aid of the Macintosh Pascal interpreter, and compiled using the TML Systems compiler. It requires a Macintosh, Macintosh Plus or Macintosh XL with 512K or more memory. The maximum number of taxa is 80; the maximum number of characters is 200. Characters can be unordered, ordered, or irreversible. Missing data are allowed. In combination with MacPaint and the Laserwriter printer MacClade produces publication quality cladograms.

Acknowledgements: David Maddison, who is already working on revisions to MacClade, deserves special thanks for continued help and advice that had much to do with the shaping of MacClade. A number of people around the MCZ, including Chuck Crumly and Hans Sues, helped with comments and the vertebrates data set (do not hold them responsible for any peculiarities in the data set). Herbert Levi and NSF provided use of a Macintosh.

USING MACCLADE

Insert the MacClade disk. When the MacClade Icon appears, move the mouse cursor over top of it and double click the mouse button (i.e., click twice in close succession). This will start MacClade. When it gives you a box with file names, open the data file you want to use by double clicking on its name or by clicking on it once and hitting the open button. I suggest you first play with MacClade using the practice data files on the disk before you make your own data file (see Practice Session, below).

MacClade then shows you parts of the data file you have opened. The taxon names are in bold face, the character information in regular type. When MacClade beeps and says "Click to continue", click the mouse. MacClade then lists the cladograms stored in the data file. Type the number of the cladogram you want to start with, and hit Return. MacClade will then make the cladogram and display it on the screen. If no cladograms are in the data file, MacClade will start with the default pectinate cladogram (cladogram 0). This pectinate cladogram could be a completely artificial cladogram for your data; it is given as default <u>only</u> because it is easy to program and provides a starting point for your own rearrangements.

Now MacClade will show you the cladogram. There are two sorts of command you can give: (1) you can click on the boxes, e.g. click Diagnos to show diagnostic information about characters, or (2) you can click on the cladogram, e.g. to perform an operation on the branch or clade on which you clicked.

Note on use of words <u>tree</u> and <u>cladogram</u>: On the screen, MacClade uses "tree" often to refer to the cladogram. This is done solely

for the sake of compactness, not as a statement about the proper distinction between trees and cladograms (Eldredge & Cracraft, 1980; Hendy & Penny, 1984).

Practice session: After starting MacClade, opening the test data file "Vertebrates". Select tree number 1. After MacClade has constructed the tree, click on "Trace Character" (Figure 1). Then type 1 and hit Return, in order to trace character 1. Note that according to the cladogram, the amnion arose three times independently (Figure 2). Then move the arrow cursor over top of the branch below crocodiles and lizards, and press down on the mouse button. The cursor should change from an arrow to a diamond. While continuing to hold the mouse button down, drag the diamond until it is over top of the branch below birds and mammals (Figure 2). Drop it (release the mouse button). The tree will rearrange to put crocs and lizards as the sister group of mammals and birds. In a similar fashion move snakes, placing it as the sister group of crocodiles (Figure 3). Now click on the branch that is the right descendant of the root, drag the diamond until it is over top of the box at left labelled "Full Swap Above" (Figure 4), and let go of the mouse button. MacClade will then try local branch swapping to find a more parsimonious cladogram (Figure 5). To look at other characters, hit "Other Character" (then type a number and hit Return), or "Next" or "Prev." or "Next 7". By looking at these other characters see if you can rearrange to an even more parsimonious tree (shortest is 20 steps).

COMMANDS GIVEN BY CLICKING BOXES

IN THE BOX AT FAR LEFT

Treelength: Clicking on this box does nothing, but the box shows how many steps of character evolution are required by the tree. The total treelength is calculated as the sum of the number of steps for the individual characters multiplied by their respective weights. It includes steps that must have occurred within polymorphic terminal taxa.

Suspend Count, Resume Count: Clicking this box turns the count of tree length on and off. Turning off the count is useful with large data sets when you want to make a lot of branch rearrangements and don't need to know tree length until you're done, for counting tree length causes delays. (If the data set is large then counting will be initially turned off and this box will read 'Begin Count'.)

Undo Branch Move, Undo Rerooting, Undo Branch Swap, Undo Get Tree (visible after a branch move, rerooting, branch swapping, or getting a tree from the data file, respectively): Undoes the last branch move, rerooting, swap, or get tree. After undoing a branch move, rerooting, swap, or get tree, clicking Undo redoes it.

Save Tree (visible when MacClade thinks the current tree on screen hasn't been saved): Saves the tree on screen to the data file. MacClade will beep and show "As:". Type your name for the tree and hit Return.

Get Tree: Gets a tree from the data file. If memory is very low (less than 10K), getting a tree from the data file may cause MacClade to crash (MacClade indicates if the memory is low).

Show Changes: displays the number of characters unambiguously changing on each branch. The darker a branch is shaded, the more characters change on it. This is useful if you want to know how many tick marks to put on a branch to indicate apparent apomorphies of the clade. If there is ambiguity in the placement of a change -- e.g., when it is equally parsimonious to suppose two independent derivations or a gain then loss -- Show Changes will not list the change except in the one case described below. To find out what characters change on a given branch, click on the branch. The characters will be listed in the box at middle right. Even though these changes along a branch can often be considered apomorphies unambiguously supporting the clade, this is not always a good interpretation. Likewise, a tree shaded darkly by Show Changes is not necessarily a good tree, for some of the unequivocal character state changes that occur on the tree might be avoided by a shorter tree. To exit from the Show Changes Mode hit "Changes off".

The only ambiguous change MacClade will indicate in Show Changes is one in the placement of a difference between the right and left descendant branches of the root. If the right descendant branch of the root is reconstructed to have state 1, while the left branch has state 0, then there is ambiguity as to whether the root had 0 changing to 1 up the right side or the root had 1 changing to 0 up the left side. That is, when the left side of the tree is characterized by 0 and the right by 1, it is ambiguous as to whether 1 is a synapomorphy distinguishing the right or 0 a synapomorphy distinguishing the left, unless the character is irreversible. Because you might be interested in knowing about these characters which might distinguish the left side of the tree from the right, MacClade indicates these by shading the root and listing them when you click on the root branch. When there are such characters MacClade gives a warning that these changes indicated for the root actually belong on either descendant branch.

Shrink (visible only when tree expanded): shrinks whole tree down to screen size (see Expand, below).

Print: Erases some of the boxes on screen and prints the tree on the Imagewriter printer. To print the screen as is, without erasing these boxes, hit shift-command-4 (see your Macintosh manual for more information on this command). (To print, the MacClade disk must contain the Imagewriter file.) (See Hidden commands below for additional information about printing.)

To save the screen to a MacPaint file for future editing and printing, hit shift-command-3. The figures of the MacClade documentation were obtained by saving the screen to a MacPaint file, and printing the MacPaint file on a Laserwriter.

All Terminal (visible only when at least one taxon has been made an ancestor): makes all taxa terminal again (see Make Ancestor, below).

Δ Data file (change data file): quits from the present data file and allows you to start another. If MacClade thinks the present tree is

not saved, it will give a warning and allows you to cancel the change data file command, or to save the tree before changing data, or to simply change data without saving the tree.

After ∆ Data file is selected, MacClade will give you a box with file names listed and the buttons Eject, Drive (both referring to the disks), Open (referring to the file selected), and Cancel. Note: hitting the Cancel button will cause MacClade to quit.

Quit: exit MacClade. If MacClade thinks the present tree is not saved, it will give a warning and allows you to cancel the quit command, or to save the tree before quitting, or to simply quit without saving the tree.

The Memory box is displayed only if the memory remaining is less than 25K or if the user has requested it (see Hidden Commands, below). It displays the maximum amount of memory remaining (you may in fact have less than this). Clicking on it does nothing.

IN THE BOX AT MIDDLE RIGHT

Trace Character, Other Character: Displays the most parsimonious (simplest) reconstructions of evolution of the selected character. After clicking the box, select the character by typing its number and hitting Return. The tree branches will be shaded to indicate what states are most parsimoniously placed on them. A legend is provided on the right of the screen showing what shade represents what state.

If more than one state can be equally parsimoniously placed on a branch, the branch is shaded by the "equivocal" pattern. If the mouse is clicked on a branch with an equivocal pattern, asterisks appear in the right hand legend to indicate what states may be equally parsimoniously placed on the branch (Fig. 4). Similarly, if a terminal taxon is polymorphic, the branch is shaded by the "polymorphic" pattern, and clicking on it makes asterisks appear showing the different states observed in the taxon. Note: if the character state legend boxes that are supposed to indicate shading are all white or strange speckled patterns, then you have likely made a mistake in coding character states (e.g., coded l [lower case L] or O [upper case o] instead of one or zero).

A box appears for each terminal taxon to show the state(s) observed in it. If the state was unknown for a taxon (coded in the data matrix as ?), no box appears.

Technical details. For unordered characters, my own formulation of the Fitch (1971) - Hartigan (1974) algorithm is used. For ordered characters, an algorithm developed by Swofford and Maddison (in prep.) is used in a slightly modified form. This algorithm represents an improvement over the Farris (1970) algorithm in that it finds all most parsimonious assignments to each branch. For irreversible characters the algorithm simply takes the lowest-value character state in a clade as the ancestral state for that clade. These algorithms have provisions for dealing with missing data, branches that have had their states fixed and taxa that have been

fixed as ancestors. Copies of the algorithms are available on request.

Note: if MacClade does not shade the basal node of an outgroup the same as would be achieved using the "outgroup algorithm" of Maddison et al (1984), this does not mean that either the outgroup algorithm or MacClade are in error, for the outgroup algorithm considers the outgroup in isolation from the ingroup, whereas MacClade would consider both when the outgroup is attached to the ingroup (see Maddison et al., p. 88).

Next (visible only when character trace is on): Normally this traces evolution of the following character. However, when the states for the next 7 characters are being displayed (see Next 7, below), then hitting Next jumps ahead 8 characters and displays the next section of the data matrix. Thus, if character 1 is displayed on the cladogram and boxes for characters 2 through 8 are shown above it, then hitting Next will trace character 9 on the cladogram and show characters 10 through 16 above it.

Next25 (visible only when character diagnostics are being displayed and there are more than 25 characters): displays diagnostics for the next 25 characters.

Prev.(previous; visible only when character trace is on): Traces evolution of preceding character, unless Next 7 is activated, in which case it jumps backwards and displays boxes for the previous set of 8 characters (see explanation for Next).

Prev25 (visible only when character diagnostics are being displayed and there are more than 25 characters): displays diagnostics for the preceding 25 characters.

TrOff (Trace Off; visible only when character trace is on): Turns off trace of character evolution.

Diagnos (Diagnostics of characters; visible only when count is on): Shows diagnostic information for characters, namely consistency index, number of steps, number of states, whether the character is unordered, ordered, reversible or irreversible, and the weight on the character. Also shows the consistency index for the cladogram. Diag off turns off the diagnostics. (If your data set has more than 25 characters, see Next25, Prev25 above.)

Next 7 (visible only when count and character trace are on): Displays boxes showing states of taxa for the seven characters following the one presently traced on the cladogram. N 7 off turns off this display.

Change Weight, Reset weight (visible only when Diagnose is on or weights have been changed): Allow you to alter the weights on characters. After you click Change Weight, MacClade will respond with "Δ wt of:" (change weight of:). Type the number of the character and hit Return. MacClade responds with "New wt:". Type the new weight, an integer from 0 to 9, and hit Return. Such changes are not written into the data file, but they are remembered until they are reset by you, or the data file is changed. Reset Weight

resets the weights to those stored in the $ line of the data file. The total treelength is calculated as the sum of the number of steps for the individual characters multiplied by their respective weights.

Reset un,ord,irr (visible only when unord, ord, irr have been changed):
Resets the specifications of unordered, ordered, irreversible for the characters to those stored in the * line of the data file.

IN THE BOX AT FAR RIGHT (visible only when character trace is on)

Char steps: This box does not issue any commands but it should be explained. Char steps shows the number of evolutionary steps required in the character by the tree, subject to the assumptions as to whether the character is ordered, irreversible, and so on (see below). The number shown is simply the number of steps; it is not multiplied by the character weight. Thus if there are characters weighted differently from 1, when you add all the character steps you may not get the tree length (see Treelength, above).

Unord, Ord, Rev, Irrev boxes: The highlighted (black) box indicates whether the character is currently assumed unordered, ordered, reversible or irreversible. To change this specification, click on the box desired. Such changes are not written into the data file, but they are remembered until they are reset by you, or the data file is changed. To reset, hit Reset un,ord,irr (see above).

Note: for two state characters, the boxes shown are "rev" and "irrev", while for multistate characters "unord", "ord" and "irrev" are shown. The reason for this is that the distinction between ordered and unordered is useful only for multistate characters. Thus binary characters can be assumed either reversible or irreversible. See further discussion under Input data files, part (3).

Unfix All (visible only when the state of a branch has been fixed): unfixes states of all branches whose states have been fixed; see below under Fix state at branch.

Show Fix (visible only when the state of a branch has been fixed): highlights those branches whose states have been fixed; see below under Fix state at branch.

HIDDEN COMMANDS

If the branches are very thin (i.e. there are more than 61 taxa on the screen), then clicking the box giving the title of the data set (in the lower left) will change the default branch shading from white to black & vice versa (when the branches are thin, black ones are easier to see but take longer for MacClade to draw). Otherwise clicking on this box does nothing. The initial branch shading will be black if the default tree number 0 is drawn, white if a tree was read from the data set.

If you click on the box at middle right that indicates the number of characters (e.g., "20 chars"), hold the mouse button down and drag the cursor over top of the title of the data set, then release the mouse button, MacClade will change to a mode which allowed me to print Figures 2-4. In this mode, MacClade will freeze after a branch is clicked on, dragged and dropped. Shift-command-4 can then be used to print the screen, shift-command-3 to save it to a MacPaint file. To unfreeze, click the mouse button. To change back to normal mode, repeat the action of clicking on the character name box, dragging then dropping on the title.

If you click on this box indicating number of characters, hold the mouse button down and drag the cursor over top of the treelength box, then release the button, MacClade will show the maximum amount of memory remaining the next time the tree is redrawn. The memory display can be turned off by repeating the action that turned it on.

COMMANDS GIVEN BY SELECTING A BRANCH ON THE CLADOGRAM

The following commands all deal with a specific branch on the cladogram. To select a branch, move the mouse cursor over the branch, click the mouse button and hold it down. The mouse cursor will then change from an arrow to a diamond, and various boxes on the screen will change to show you some of the commands that are possible. Also, on the left a box will indicate how many taxa are in the clade above the branch selected. Then drag the diamond, still holding down the mouse button, over top of the appropriate box or another branch, then let go of the button. If the diamond is dropped on a box then that command will be effected; if dropped on another branch then a branch move may occur.

Branch move: A branch can be picked up and moved to another part of the tree. Click on the branch, drag it over to the branch to which you want it attached, and drop it. In some circumstances MacClade will refuse to move a branch because it would be inappropriate (for instance if the terminal taxon has been made an ancestor).

To issue the following commands, select a branch and drop it over the appropiate box.

Full Swap above and Single swap above: Dropping a branch on these boxes performs local branch swapping in the clade above the branch selected (to select the whole tree, grab the root). The tree below the clade is left untouched. The branch swapping algorithm visits each branch in the clade, and sees if a shorter tree would be

obtained by rearranging relationships between its sister branch and its two descendant branches. The Single Swap finds the single swap throughout the clade that shortens the tree the most. The Full Swap does this repeatedly until it can no longer find a local rearrangement that improves length. To help you track where the swapping is taking place, each time a branch is swapped, the branch moved is shaded black (or white if already black). In each case the black branch has been moved onto its ancestor's sister branch. While this algorithm may help to fine-tune a cladogram, do not rely on it to find the shortest tree! Once I had a data matrix whose shortest tree was 36 steps, yet the local branch swapping could find no tree shorter than 72 steps, because the input tree required more than just local rearrangements. The Swap Above boxes are visible only when the branch selected is suitably far down from the terminal taxa.

Single Swap and Full Swap differ in one other respect. If no rearrangements will improve length, and yet there is a rearrangement which will keep the same length, Full Swap will make no change in the tree but Single Swap will make that equally parsimonious rearrangement. Thus you can use Single Swap to search for equally parsimonious trees.

Note that because the swapping algorithms visit the branches on the left side of the tree first, MacClade will preferentially make swaps on left side.

During Full Swap, to temporarily suspend swapping place the mouse cursor in the Suspend box (you don't have to press the button down); to abort swapping and show the tree obtained so far by swapping, put the cursor in the Abort box.

Full swap below and single swap below: These do the same as Swap above, except the swapping is done in the area of the tree below the branch selected, while the clade above the branch is left untouched. The Swap Below boxes are visible only when the branch selected is suitably far up from the root.

Expand (visible only when a non-terminal branch is selected): This expands the clade above the selected branch so as to fill the whole screen (Figs. 5, 6). This is especially useful when there are many taxa. Shrink will shrink the tree back to screen size.

Rotate: dropping on this box rotates the branch so that the right descendant clade of the branch becomes the left clade & vice versa. Of course, this does not change treelength or character evolution reconstruction. It is a feature for aesthetic purposes.

Make ancestor (visible only when terminal taxon is selected): This forces the selected terminal taxon to become an ancestor. The taxon is put in the place of the branch that used to be its immediate ancestor. The name of the taxon remains on top, but instead of a full branch the taxon has only a line pointing to its new position as ancestor. All of the character states of the taxon are forced down onto this ancestral branch. This is useful for paleontologists, because they can see the number of extra steps in character evolution required by supposing a fossil is an ancestor. It is also useful for neontologists to determine if a species might

be paraphyletic according to the cladogram: if the tree length goes up when it is made an ancestor, then the species has autapomorphies. Show Changes (see above) can also indicate autapomorphies.

Even though a taxon made ancestral has only a line coming down from the taxon's name instead of a full terminal branch, this line can still be grabbed with the mouse, though only one thing can be done with it. If dropped over Make Terminal, this taxon will be released from its ancestral position and become terminal once more. Taxa made into ancestors can be released all together by clicking on All Terminal.

A taxon fixed as ancestor remains so fixed even when another character is traced. Ancestors are automatically made terminal when branch swapping is invoked or during rerooting and certain branch moves. The fixing as ancestor is not stored when the tree is saved to the file, nor is it remembered on Undo Get Tree.

Warning: If a terminal taxon polymorphic for some character is forced down as an ancestor, the number of steps required is likely to be underestimated. A warning to this effect is given by MacClade. For example, when a terminal taxon polymorphic for 0,1 is forced down as ancestor into a region of the cladogram with all state 0 surrounding, an extra step should be counted (gain then loss of 1) but isn't. The shading of branches indicating character evolution should be correct, however.

Reroot: dropping a branch in this box reroots at this branch. If the whole tree is on the screen, then the whole tree is rerooted at the branch selected. If one clade has been expanded to fill the screen, then rerooting only reroots the clade shown; the rest of the tree below the clade remains untouched. This can be useful to try different rootings of the ingroup. To do this, expand the ingroup and reroot at various branches.

Fix state at branch (applicable only when a character is being traced): When a branch is selected and then dropped over top of one of the shaded boxes in the right hand legend, which indicate the shading used for a character state, the branch is forced to have that state regardless of what it costs in terms of number of steps. The surrounding branches are re-shaded as is appropriate, and the number of the steps is recalculated, given the fixing of the branch.

This is useful when you want to see how much it would cost in number of steps to have an alternative state ancestral for a particular clade. It is also useful if the initial shading of the tree had equivocal areas (where more than one state could have been placed at a branch with equal parsimony). One can then see one of the resolutions of this ambiguity by fixing the branch to one of these equally parsimonious alternatives (for instance, the alternative least favorable to any evolutionary hypothesis you may be testing) after which MacClade will trace character evolution in the rest of the tree given this fixing.

If an already-fixed branch is selected, a box on the right will say Unfix. Dropping on this box will free that branch to be shaded without fixing. To unfix all branches at once, hit Unfix All. To see which branches on the cladogram have had their states fixed, hit

the Show Fix box; the fixed branches will change shading temporarily.

The state of a fixed branch is unfixed if another branch is moved onto it, or if a different character is traced, or if the character trace is turned off (for instance for branch swapping or show changes), or if Get Tree is called.

Warning: if you fix the state of a terminal branch to a state other than that observed for its terminal taxon, you are really asking "what if the taxon really had this state?" Although one might count an extra step between this state and the state actually observed for the taxon, MacClade does not do so. MacClade does not write such changes in the states of taxa into the data file. As stated above, such fixings are unfixed when another character is traced or trace is turned off.

NOTES REGARDING PERMANENCE OF CHANGES

MacClade might be viewed as a cladogram and character editor, in that it allows you to manipulate the cladogram and character evolution on screen. Some of the changes you make are more permanent than others. The fixing of a state at a branch is rather temporary since it is lost if a different character is traced and can't be saved to the data file. The fixing of a taxon as ancestor is a bit more resistant to loss but likewise it can't be saved to the data file. For more information on when these fixings are lost see the instructions above regarding them. Changes in character weights and the unordered/ordered/irreversible specifications are remembered by MacClade throughout the use of one data file, unless reset, but can't be saved to the data file. Changes in the tree structure are remembered until further changes are made, and the new trees can be saved to the data file. After you make a change in tree structure, MacClade remembers the previous tree, so you can undo the change and recover it.

INPUT DATA FILES

For MacClade to be able to read a data file, the taxon names, coding of character states, and so on, must follow the format described below. There are two ways to make a data file with this format: (1) make it directly with a word processing program such as MacWrite or Microsoft Word and follow the formatting rules below, or (2) make the data matrix with the spreadsheet programs Multiplan or Excel and allow MacClade to convert the matrix to the MacClade data file format. Most of the following paragraphs describe how to make and format a data file using a word processor; instructions for use of Multiplan and Excel are given at the end of the section.

USING A WORD PROCESSING PROGRAM

If you use MacWrite or Microsoft Word to write the data files, make sure you click the Text Only box under Save As. If the files are not saved as Text Only, MacClade will not even recognize that they exist. It will be useful to use the Monaco 9 type face when editing the files because this typeface allows the characters to line up. The desk accessory MockWrite and the general text editor Edit are both useful ways to edit, for they naturally save all files as text only and use Monaco 9.

Warning: I have put few error handling routines in MacClade. If you have incorrectly made the data file, MacClade might either crash or give you strange results. A discussion of possible errors & crashes is given below.

MacClade reads data files in a fairly flexible way, allowing you to avoid writing format statements and to avoid specifying number of taxa and characters.

There are several example files on the MacClade disk, including 20x22 test data, 80x200 test data, Pell. chromosomes and Vertebrates. The 20x22 and 80x200 test data are nonsense data sets. Indeed, to get 80 taxa for the 80x200 data set, I duplicated 16

times the lines for 5 taxa. The Pell. chromosomes is a real data set, though the morphological data on which the cladogram was based are not included. The Vertebrates data set is example 2 below and was used in the practice session above.

Example 1. A very simple data file with 4 taxa and 3 characters:

```
This is the title
Homo sapiens  011
Drosophila    102
Rattus rattus 100
E. coli       011
*
$
```

Examples 1 and 2 show two data files. You might think of the data file as consisting of four parts:
(1) Introduction.
(2) The data matrix.
(3) The specification of character assumptions (* and $ lines).
(4) The storage area for trees (optional).

General rules: Blank lines are allowed as long as they have no characters (not even spaces) other than the carriage return. Except where noted, lines can be up to 255 characters long. To get male and female symbols in taxon names, character names, and so on, type option m and option f respectively. These usually appear as and , but MacClade displays them as and . Comment lines are allowed after the title and before the * line. A comment line must begin with a double quote character ("); the line is ignored by MacClade. In example 2, a comment line that has character numbers "123456789.123" is used to aid reading the data matrix. Comment lines in the data matrix can be used to separate off the taxa in blocks of 5 as in example 2.

(1). Introduction. The first line of the data file must be the title.

Optional: Following the title you can add, if you want, lines that specify the names of characters and the names of character states. In example 2, character 1 is called "amnion" and character 7 "temp.fenestrae". Begin these lines with the number of the character (if you start the line with a space, MacClade will not recognize it properly). After the number put a period and a space. If you want to give names to the character states as well, put a colon after the character name. Then, for every state from 0 through the highest state number, put a space, statename, semicolon. Thus, these lines take the form "#. character name: state 0 name; state 1 name; state 2 name;". In example 2, character 12 has state 0 "ossified" and state 1 "not". Each of these lines with character and state names must contain no more than 128 characters. Any text occurring after the last semicolon in each of these lines is ignored by MacClade and can be used for comments (e.g., "(archosaurs)" at the end of the line for character 12 of example 2.

(2) The data matrix. The first line after the title that begins neither with a number or " is interpreted as the first line of the data matrix.

In the data matrix, you can make the taxon names as long as you like, up to 16 characters, as long as you are consistent, i.e. the

character information must start in the same column for each taxon. Tabs are not allowed in the data matrix!! Note: the name of the first taxon cannot include any numbers (e.g., "Genus#2 n.sp." is illegal).

After the name of the taxon, the character states for the first through the last characters are specified by numbers (integers between 0 and 9). The character state for a taxon must be aligned with that of the other taxa, i.e. if character 3 appears in the 15th column for one taxon, it must appear there for all taxa. No blank spaces are allowed between the first and last characters on a line, or after the last character. Specify missing data by ?.

If you want to have characters polarized, outgroups must be entered into the data matrix explicitly. State 0 is not automatically assumed to be ancestral.

To indicate that a taxon is polymorphic for a character, add a line just below the taxon, beginning with a &. In the column corresponding to the character, place the number for the second state that the taxon has. Additional & lines can be added if the taxon has more than two states for a character. In example 2, mammals are coded as polymorphic for character 2, having states 1 and 2. It may seem strange that mammals are coded as polymorphic for appendages. A coding of polymorphic for a species may mean that several states coexist in its populations, but clearly that is not what was meant since Mammalia is a class. For Mammalia the coding means that the taxon includes some members with legs only and others with legs and wings, and since we are ignorant of which state is ancestral for the taxon (for the purpose of the analysis), we allow either state to be ancestral by coding the taxon as polymorphic. MacClade prohibits polymorphisms to be retained in the branches below the terminal taxa, except if a polymorphic taxon is forced down to ancestral position (see Make Ancestor, above).

(3) Specification of assumptions about characters. Immediately following the data matrix must come one line beginning with *. This marks the end of the data file. Immediately following the * line must come a line beginning with $.

Optional: If you want, you can use the * line to indicate whether the character is unordered, ordered or irreversible. In an unordered character, a change from any state to any other state is counted as one step, whether it is from 0 to 1 or 0 to 8 or 7 to 4. In an ordered character, a change from 0 to 1 is one step, 0 to 8 is eight steps, 7 to 4 is three steps. An irreversible character is like an ordered one but the character state cannot decrease. To indicate your choice, place 'u' (unordered), 'o' (ordered), or 'i' (irreversible) in the * line in the column corresponding to the character. If you make no indication, unordered is assumed. With two-state characters there is no useful distinction between ordered and unordered (see Meacham, 1984). Thus, the only distinction MacClade makes for two-state characters is between reversible (0->1 and 1->0 transitions allowed) and irreversible (only 0->1 transitions allowed). Because of this, MacClade will treat 'u' and 'o' specifications as equivalent for two-state characters and will label characters so specified as "rev" in the legend or "r" in the

character diagnostics. If you make no indication for a two-state character, reversible is assumed. These specifications can be changed while MacClade is running.

Optional: If you want, you can use the $ line to indicate the weight of characters. The total treelength is calculated as the sum of the number of steps for the individual characters multiplied by their respective weights. The weight must be a number from 0 to 9, and is placed in the $ line in the column corresponding to the character. Weight 1 is assumed unless otherwise specified. Weights can be changed while MacClade is running.

(4) Tree data. (Optional.) You don't need to put a tree into your data file initially, but you can if you want. Each tree takes up two lines in the file: (a) Title of tree; (b) specifications for tree. Line (b) must immediately follow line (a). Only dichotomous trees are accepted by MacClade. Use parenthesis notation for trees, as in Example 2, with a semicolon at the end of the line. In example 2 the first tree ("Ridiculous") was typed in by me, the second tree ("Same as Ridic.") is the same tree, only saved in the file by MacClade using its slightly more compact form. The taxon numbers come from the order of the taxa in the data matrix. Thus the "Ridiculous" tree has taxa 4 (lungfish) and 10 (snakes) as sister groups. If you rearrange the order of taxa in your data matrix without editing the tree specifications you will end up with some unexpected trees!! You don't need to add right hand parentheses; they are redundant since MacClade accepts only dichotomous trees. Indeed if you have more than 50 taxa you shouldn't use right parens since they may make the input too long. If you have more than 60 taxa you shouldn't put spaces or commas after two-digit taxon numbers, to save further on space. If you input a tree specification that is incorrectly formulated or longer than 255 characters, MacClade may crash or may show you a tree lacking some of the taxa!! You may be better off starting with the default pectinate tree that MacClade supplies, rearranging on screen to the desired tree, then saving it to your file. Hint: when doing many rearrangements on a large tree, expand small clades and rearrange within these as much as possible because it takes less time for MacClade to draw a small clade. When MacClade saves trees it will save them at the end of your data file.

Polychotomous cladograms are not yet accepted by MacClade. In general, the character evolution and length of a polychotomous cladogram can be calculated exactly only if one assumes the polychotomy to be real (i.e. multiple speciation) and not an uncertainty. If one views the polychotomy as an uncertainty in resolution (as is the usual interpretation), and the polychotomy is left unresolved, then in general character evolution and tree lengths can be only approximately determined. Because of these difficulties, the benefit of allowing polychotomies seemed less than the cost. Allowing polychotomies is planned for subsequent versions of MacClade, however. For now you can try inputting one of the resolutions of the polychotomy and then rearrange on screen to try other resolutions. This indeed may be the most informative way to learn about the polychotomy, as long as there aren't too many resolutions to try.

Comment lines are not allowed in the tree data section.

USING MULTIPLAN OR EXCEL

MacClade automatically converts spreadsheet files made using the Microsoft spreadsheet programs Multiplan and Excel into MacClade data files, thus allowing you to use the special matrix-handling capabilities of these programs to conveniently edit your data. These spreadsheet programs allow you to insert and delete columns, and put character names at the tops of columns.

For MacClade to be able to read the Multiplan or Excel file, it must be saved as a SYLK file (choose SYLK under Save As when you make your file). When MacClade starts to read a SYLK file, it recognizes it as such and translates it into a MacClade data file with the format described above. This newly made MacClade data file has the same name as the original SYLK file but with ".mcl" added. After translating the SYLK file and storing the new MacClade data file on disk, MacClade then opens the new MacClade data file to begin your analysis.

When making the data file in Multiplan or Excel these rules must be followed: in the first cell (first column, first row) the title of the data must appear; in the first column the taxon names must appear; in the first row the character names must appear. Thus, each row (below the first), corresponds to a single taxon, and each column (after the first) corresponds to a character. The first taxon name must not include a number and the character names must not include colons (:). The data themselves must be input as integers for observed character states or ? for missing data. Single digit integers (from 0-9) indicate a monomorphic taxon; multiple digit integers indicate polymorphisms. Thus 3 indicates the taxon has just state 3; 10 indicates it has states 1 and 0, while 201 indicates it has states 2, 0 and 1. Polymorphisms of more than 5 states are not allowed: 10736 is allowed (5 digits) but 216304 (6 digits) is not. Do not leave any cells blank! If there are 50 taxa and 30 characters, then every cell must be filled in the upper left 50x30 block of the spreadsheet. The SYLK conversion facility in MacClade may not work if you use fancy formatting or cell protection in your spreadsheet, so be simple. You may use Freeze Titles.

There are some disadvantages using Multiplan and Excel: character state names, tree data, character weights, and ordered/undordered etc. specifications cannot be included in the spreadsheet. You may find it most convenient to construct the initial data matrix in Multiplan or Excel, translate it using MacClade, then edit it further using a word processor to add the character state names and so on. Note that if you make a Multiplan file named, for instance, Example, then use MacClade to translate it into a MacClade data file Example.mcl, then you modify Example and translate it once again into Example.mcl, the first Example.mcl will be erased and replaced by the second one. Thus, if you have worked with the first Example.mcl, saved trees in it, edited it using a word processor, and so forth, you may want to change its name to something else to avoid having it overwritten by the second Example.mcl.

ERRORS & CRASHES

MacClade gets stuck when accessing the data file: you may have forgotten the * or $ lines.

MacClade crashes with ID=02 error when it first starts and is trying to present you a list of file names to open: this happens sometimes when you have just finished using an application on another disk; why, I don't know. Try resetting Mac or turning it off and on, and begin again.

MacClade crashes spectacularly while reading data matrix: there may be blank lines that have spaces on them (blank lines are allowed only if they have nothing but the carriage return). Or, you may have forgotten the period after the character number when specifying the character name. Or, the file read may not be a MacClade data file.

MacClade gives a warning that characters are miscoded: you may have typed an l (lower case L) or O (upper case o) instead of one or zero; you may have put tabs in the data matrix; the columns of the data matrix may not be lined up; you may have put spaces beyond the last character on the first line of the data matrix; the first taxon name may have included a number. MacClade will not necessarily warn you about all possible miscodings.

MacClade gives a warning that some taxa are missing from the tree data: you may have miscounted parentheses or forgotten to include the number for a taxon in the tree specification, perhaps because a new taxon was just added.

ABOUT BUGS

The algorithms used by MacClade to count steps (Char steps and tree length) and reconstruct character evolution have been thoroughly checked and tested on a number of data sets, though some bugs may remain in some of the more complicated cases (for instance when taxa are fixed as ancestors, states at branches are fixed and there are missing data). More likely are bugs in screen-display and giving commands. I would much appreciate hearing of any bugs you find. Please record as much information concerning the context of any bugs. If the system crashes, please record the ID number of the error.

PLANS FOR FUTURE

David Maddison and myself are already working on revisions to MacClade. Later releases will hopefully have:

1. Ability to edit the data file without exiting MacClade.
2. Additonal tree and character manipulation features.
3. Exclusion of selected taxa.
4. Use of the standard Macintosh user interfaces, including menus, dialog boxes, and desk accessories.
5. Larger data matrices. Our calculations suggest that the memory of a Mac Plus could allow at least 200 taxa x 1000 characters (though one would have to be very patient in using such a matrix!). MacClade presently limits number of taxa to 80 because of screen size and resolution. The taxon number could be increased by having better techniques to view only part of a large cladogram at any time.
6. Ability to share data files with PAUP and PHYLIP, which may be available on the Macintosh in the near future.
7. Some of your suggestions!

REFERENCES

Eldredge, N. & J. Cracraft 1980. Phylogenetic Patterns and the Evolutionary Process. Columbia U. Press, New York.

Farris, J.S. 1970. Methods for computing Wagner trees. Syst. Zool. 19: 83-92.

Fitch, W.M. 1971. Toward defining the course of evolution: Minimum change for a specific tree topology. Syst. Zool. 20: 406-416.

Hartigan, J.A. 1973. Minimum mutation fits to a given tree. Biometrics 29: 53-65.

Hendy, M.D. & D. Penny 1984. Cladograms should be called trees. Syst. Zool. 33: 245-247.

Maddison, W., M. Donoghue, & D. Maddison 1984. Outgroup analysis and parsimony. Syst. Zool. 33: 83-103.

Meacham, C.A. 1984. The role of hypothesized direction of characters in the estimation of evolutionary history. Taxon 33: 26-38.

Example 2. A more complex datafile with 10 taxa and 13 characters. This is one of the practice data sets on the disk.

```
Some Vertebrates
"
"character & state names ============
1. amnion: absent; present;
2. appendages: fins; legs only; legs&wings; (tetrapods)
"        (Note that we know enough about phylogenetic structure of
mammals
"         that we could code them as "legs only" but let's be naive.)
3. body covering: dermal scales; smooth; epid.scales; feathers; hair;
4. thermoreg.: poikilotherm; homeotherm;   (birds & mammals)
5. int.nostrils: absent; present;          (sarcopterygians)
6. atrial septum: absent; present;         (sarcopterygians)
7. temp.fenestrae: none; one; two;         (diapsids)
8. hemipenes: absent; present;             (squamates)
9. suspensorium: non-strep; streptosylous; (squamates)
10. gizzard: absent; present;              (archosaurs)
11. antorb fenes: absent; present;         (archosaurs)
12. lat.sphen: ossified; not;              (archosaurs)
13. teeth: not ped.; pedicillate;          (lissamphibians)
"====================================
"==Data=======123456789.123
"0===========+=========1===
rayfinned fish00000000?0010
frogs         0110110000011
turtles       1120110000010
lungfish      00001100?0010
salamanders   0110110000011
"5===========+=========1===
crocodiles    1120112001100
lizards       1120112110010
birds         1231112001100
mammals       1141111000010
&              2
snakes        1?20112110010
"10==========+=========1===
"=============123456789.123
*              o    oo
$
Ridiculous
(1,(((4,10)((6,7),5)),(2,(3,(8,9)))));

Same as Ridic.
(1 (((4 10((6 7 5 (2 (3 (8 9 ;
```