BOOK REVIEW

MacClade Version 3: Analysis of Phylogeny and Character Evolution. By Wayne P. Madison and David R. Madison. Sinauer Associates, Sunderland, MA, 1992. xi + 398 pp. Paperback + MacClade Computer Package for the Macintosh, Version 3.03, ISBN 0-87893-890-1, $75 with software, $40 book only.

The process of transforming raw molecular data into phylogenetic trees is a treacherous one. First, one must go through a long series of steps, including multiple alignment, correction of the alignment by what is euphemistically called "visual inspection," calculation of genetic distances, identification of informative sites, the phylogenetic reconstruction itself, and the determination of the reliability of the particular clusterings in the tree. Each such step requires the making of several difficult decisions and choosing among sometimes incompatible alternatives. Should one use DNA or translate it into protein? Or, maybe the alignment should be performed at the level of the amino acid, but the calculations should be made on the DNA? What particular method of correction for multiple substitution should be used? Is a metric system of phylogenetic reconstruction appropriate or should the character states be used directly? And if one chooses the former alternative, should one use the neighbor joining method or Sattath and Tversky's method? And how should one decide how much confidence should be placed on the topology of the tree? Should the result be bootstrapped because *cosi fan tutte* or should one calculate the variance of the lengths of the internal branches?

And these are the easy questions. One must also decide on the value of such parameters as gap penalties, the size of the window, the substitution matrix, and the weights for transitions versus transversions. And if these indignities are not sufficient, one may also need to root the tree, calculate branch lengths, translate the number of substitution into times of divergence, and decide on a method to draw the tree so as to convey a clear picture of the evolutionary history of the sequences in question.

Most bewildered molecular biologists do not perform these tasks manually, but rely on various computer packages and programs. Essentially, from the moment the sequencing gels are read, most biologists perform a series of rites and rituals in a prescribed order, and the end result is printed and sent to a journal. From EMBLSEARCH or CDC to STRIDER to CLUSTALV to PHYLIP or PAUP through DNAPARS, PROTPARS, LWL89, and NJ, the sequences enter the hard disk and trees exit through the personal laser printer. And since no one knows the right values for gap penalties or the meaning of window size, the researcher thanks his lucky star for the default options and presses RETURN. This is GiGo (garbage in, garbage out) in its purest form. In the process, most of us inevitably lose touch with the data, and the phylogenetic trees become little more than obligatory icons to adorn a paper that is not in fact concerned with evolution at all.

MacClade is aimed at putting the researcher in touch with his data again. It is not designed to look for the most parsimonious phylogenetic tree, nor to indicate how much confidence we should place on a particular clade. Rather, it is meant to allow one to "play" with the data and see what happens if certain changes are made for intuitive or arbitrary reasons. How much longer will a tree be if this branch is moved to that place? Or, will the tree be more understandable to the reader if two OTUs originating from the same node are rotated about their axes? It can also be used to combine different sets of data, for instance to use a molecular tree in order to determine the dynamics of evolution of morphological data.

The book is divided into three parts. The first part contains an introduction and a tutorial. The second part is a theoretical essay on cladistics and parsimony methods. It bears the title "Phylogenetic Theory" but words like "neighbor joining," "maximum likelihood," and "genetic distances" are treated as taboos, neither touched nor mentioned. Part two of the book can be purchased separately if one is so inclined. Part three of the book is a manual that is indispensable in operating *MacClade*. Although *MacClade*, like many other programs written for the Macintosh, is a "friendly" program, it is not that friendly or self-explanatory that it can be approached without a very long set of instructions.

Many of the features of *MacClade* are concerned with morphological characters and they should not concern too much the readers of *Molecular Phylogenetics and Evolution*. I assume that most molecular evolutionists will start using this program by importing a tree obtained by a particular method of phylogenetic reconstruction into *MacClade*. *MacClade* accepts data and trees from PAUP and PHYLIP, and since neighbor joining trees can be produced by CLUSTALV and exported into PHYLIP, one can also use these trees by passing the data through an intermediary. Data are

entered into *MacClade*'s Data Editor, and there they can be manipulated, renamed, subdivided, merged, deleted, and filtered. The options are spectacular. Fonts can be changed, pictures can be used instead of taxonomic names, and pictorial icons can be used instead of character-states, to yield very pleasant-looking figures. Unfortunately, *MacClade* does not deal properly with gaps, a feature it shares in common with PAUP, but we are promised that "future versions of MacClade will allow better ways to deal with gaps."

The main part of *MacClade* is contained in the Tree Window, in which the relationship between phylogenetic tree and the data is explored. Trees are manipulated by using tools in a Tools menu, and here it is that the real fun begins. One can move branches and clades around and see what happens to the length of the tree; one can reroot the tree, turn a binary tree into a concensus tree by creating polytomies, remove branches, and rotate branches for more convenient display. Most importantly one can start tracing character evolution along the branches, and see what happens at position 235. A parallel substitution from A to C has occurred in geese and vampire bats! Well, we may

have discovered a case of convergence due to horizontal gene transfer, and this possibility should be investigated quickly.

Finally, *MacClade* computes consistency and retention indices, charts trees and character statistics, compares different trees (tree 1 is shorter than tree 2 by 23 substitutions, but all the substitutions occur on the branch leading to *Drosophila*), computes the number of steps required at each position and along each branch, and everything is ready to be printed in publishable form.

MacClade takes the monotony out of phylogenetic reconstructions. It is a lively plaything that gives delight, and in the process may help the thoroughly bootstrapped researcher gain some human insight that he might have otherwise missed.

DAN GRAUR
Department of Zoology
George S. Wise Faculty of Life Sciences
Tel Aviv University
Ramat Aviv 69978
Israel