

REVIEW

Cladistics (1993) 9:351–353

COMPONENT, Version 2.0.—Roderic D. M. Page. 1993. The Natural History Museum, London. User's license: £40.00 (includes manual); existing users of COMPONENT 1.5 may upgrade for £30.00. Runs under Microsoft Windows 3.0 or later, and requires an IBM PC or 100% compatible with Intel 80286 or higher processor, 2 Mb RAM, and a hard disk.

COMPONENT 2.0 is a computer program for comparing phylogenetic trees, an activity central to several subdisciplines of systematics, notably historical biogeographic and host–parasite studies. It is *not* a program for inferring trees from data; instead, pre-existing trees must be provided to the program. COMPONENT is capable of a variety of methods of tree comparison, including computing consensus trees, calculating the similarity between pairs of trees, and mapping one tree onto another. Additionally, COMPONENT offers a number of other options not involving the direct comparison of trees. A number of these capabilities are not currently available in other phylogenetic programs (e.g. PAUP, PHYLIP, Hennig86), and in this sense, COMPONENT is complementary to these programs. Version 2.0 offers several options not available in Version 1.5, including additional tree comparison and consensus methods, tree statistics and tree mapping. Additionally, COMPONENT 2.0 employs the NEXUS file format introduced by PAUP 3.0 (Swofford, 1990) and MacClade 3.0 (Maddison and Maddison, 1992), making it compatible with those programs. COMPONENT 2.0 operates under Microsoft Windows, which employs an interface very similar to the Macintosh interface, with a mouse-controlled pointer, icons, menus, scroll bars, etc. Mac users will adapt to COMPONENT very quickly.

Once pre-existing trees are input, several tree comparison methods are available. These fall into three categories: calculating consensus trees, calculating a measure of similarity/dissimilarity between two trees, and mapping one tree onto another. COMPONENT allows the user to choose between Adams, strict, majority rule, semi-strict, "Nelson" and agreement subtree consensus methods. The latter two methods are found in no other phylogenetic program. The Nelson consensus method contained in COMPONENT is the compatibility method formalized by Page (1989), not the strict consensus method with which it is often equated. The agreement subtree method finds the largest identical subtree shared by two trees.

COMPONENT 2.0 offers several options for calculating the similarity/dissimilarity between two trees. These are the partition, quartet, triplet, nearest neighbor interchange and agreement subtree metrics. Of these methods, only the partition metric is available in another generally available phylogenetic program (PAUP 3.0: Swofford, 1990). COMPONENT allows one to use the tree comparison metrics to compare all trees within one profile or between two profiles. Besides outputting the individual values to the display buffer, the program also displays a histogram representing the number of pairwise tree comparisons for each value of the metric. This very useful feature readily allows one to calculate the average similarity for a profile of trees. The probability of a value for a tree comparison metric can be estimated using COMPONENT's random tree generator. This useful feature generates random trees based on one of three assumptions: every labelled tree is equiprobable,

every unlabelled tree is equiprobable or every labelled dendrogram (nodes ordered through time) is equiprobable. [These are the tree distributions discussed by Simberloff et al. (1981).] The latter distribution is simply the familiar "Markovian" tree distribution of Harding (1971).

Perhaps the feature of COMPONENT that will be of greatest interest to users is its tree mapping capability. Tree mapping is a procedure involving two incongruent trees wherein one tree (the "associate" tree) is modified to fit the other (the "host" tree), which is left unmodified. This procedure is useful in several situations including mapping a parasite tree onto a host tree, a species tree onto an area tree or a gene tree onto a species tree. Currently, two tree mapping methods exist. The first, Brooks parsimony analysis (BPA: Brooks and McLennan, 1991), converts the associate tree into a set of additive binary characters and then maps them onto the host tree by parsimony. Homoplastic changes are interpreted as dispersals and extinctions.

The second procedure, "tree reconciliation", developed by Page (1990), is the method offered by COMPONENT. Tree reconciliation works by assuming as many duplications and losses (extinctions, unsampled taxa) of the associates as are necessary to make the associate tree exactly fit the relationships of the host tree. The fit of the host tree to the reconciled associate tree can be quantified by three different measures: number of duplications, number of terminal nodes added and number of independent losses. The similarity of the host and associate trees can be tested statistically by using COMPONENT's random tree generator to generate a large number of random trees. The original associate tree is then reconciled with each random tree, resulting in an estimated probability distribution for each of the three measures of fit.

As acknowledged by Page (1993), tree reconciliation often results in unrealistic reconciled associate trees. If one or more associates have dispersed, the reconciled associate tree may contain excessive numbers of duplications and losses. Page (1993) recommends deleting an individual associate taxon and then reconciling the host and associate trees. If the deletion significantly reduces one of the three measures of fit, it suggests that the associate dispersed. A menu command allows this to be done automatically for all associates. This *post hoc* method can significantly improve the reconciled tree, but nonetheless is rather inefficient. Page (pers. comm.), however, has recently developed an algorithm that allows duplications, losses and dispersals to be considered simultaneously in finding a reconciled tree. This method works by finding the reconciled associate tree that maximizes the number of "co-speciation" (the reader will remember that the method is not restricted to host-parasite studies) events. In Page's usage, a co-speciation event is a speciation event of the host that results in speciation of the associate. The algorithm is contained in a separate program called "TreeMap", which will soon be available by request from Page. Page (pers. comm.) plans to incorporate this method into COMPONENT in the future.

Several useful methods not directly involving tree comparison are available in COMPONENT. Foremost among these is the random tree generator, which allows one to statistically test values for the comparison metrics as well the fit of reconciled trees. COMPONENT also contains an option for enumerating all possible labelled or unlabelled trees for up to a certain number of terminal nodes. Another option calculates two statistics for input trees, the Sourdis number and Harding's topology number. These numbers are simply standardized indices for labelled and un-

labelled, binary trees, respectively. Knowing the Sourdis number for a given labelled tree, for example, allows one to recover that tree. These statistics are especially useful when it is necessary to store a large number of trees in a limited amount of space. Lastly, COMPONENT has options for modifying as well as printing trees. Tree modification is accomplished by invoking the "tree editor", allowing the user to alter a tree in different ways, including deleting terminal nodes, collapsing interior branches, moving branches, etc. The tree printing abilities of COMPONENT are not especially sophisticated, but tree files can be saved as graphics files for manipulation/printing with a graphics program.

Generally, COMPONENT's operations are quite fast. I experimented with running different options on up to 1000 trees containing up to 20 taxa. With the exception of the quartet, triplet and agreement subtree metrics, all operations took less than one hour.

Besides the problems discussed above with the tree reconciliation method, my other substantive criticism of COMPONENT concerns the options for statistical testing. Currently, the program contains no option for testing the significance of the similarity of a profile of trees considered as a whole. Such a task would be useful, for example, in testing a vicariance biogeographic hypothesis. Also, the statistical testing procedure for the fit of two reconciled trees is cumbersome. It requires generating a set of random trees, reconciling the associate tree to these random trees, and then counting the number of reconciliations requiring as many or more of the modifications required by the reconciliation being tested. It would be easier if COMPONENT contained an option that automatically performed these steps. However, COMPONENT is a "work in progress" and it is likely that future versions will remedy these problems.

Overall, COMPONENT is very user-friendly and takes a minimum amount of time to master. Furthermore, the manual is easy to read and ends with a valuable chapter containing worked examples. COMPONENT contains a number of tree comparison methods not available in other phylogenetic programs. Any systematist or evolutionary biologist with a serious interest in modern phylogenetic techniques should become a licensed user.—**Joseph B. Slowinski, Museum of Natural Science, Louisiana State University, Baton Rouge, Louisiana 70803, U.S.A.**

REFERENCES

- BROOKS, D. R. AND D. A. MCLENNAN. 1991. *Phylogeny. Ecology and Behavior*. Chicago Univ. Press, Chicago.
- HARDING, E. F. 1971. The probabilities of rooted tree-shapes generated by random bifurcation. *Adv. Appl. Prob.* 3: 44–77
- MADDISON, W. P. AND D. R. MADDISON. 1992. *MacClade: Analysis of Phylogeny and Character Evolution, Version 3.0*. Sinauer Associates, Sunderland.
- PAGE, R. D. M. 1989. Comments on component-compatibility in historical biogeography. *Cladistics* 5: 167–182.
- PAGE, R. D. M. 1990. Component analysis: a valiant failure? *Cladistics* 6: 119–136.
- PAGE, R. D. M. 1993. *User's manual for COMPONENT, Version 2.0*. The Natural History Museum, London.
- SIMBERLOFF, D., K. L. HECK, E. D. MCCOY AND E. F. CONNOR. 1981. There have been no statistical tests of cladistic biogeographical hypotheses. *In*: G. Nelson and D. E. Rosen (eds). *Vicariance Biogeography: a Critique*. Columbia Univ. Press, New York, pp. 40–63.
- SWOFFORD, D. L. 1990. *PAUP: Phylogenetic Analysis Using Parsimony, Version 3.0*. Illinois Natural History Survey, Champaign.