

Reliable phylogenetic resolution of morphological data can be better than that of molecular data

Richard H. Zander

Missouri Botanical Garden, P.O. Box 299, St. Louis, Missouri 63166, U.S.A. richard.zander@mobot.org

The use of nonparametric statistics and null hypotheses (Kendall & al., 1987) are powerful tools in science, and are essential in evaluating the question “What is the chance that these data are due to chance alone?” They have been widely used in systematics only recently with the rise of techniques in statistical phylogenetics that can deal with obscure and conflicting data (Avice, 1994: 314; Philippe & al., 1996; Sites & al., 1996; Seberg & al., 1997).

A major problem of cladistic analysis in the past 30 years has been lack of a means to measure the degree of support for contrary arrangements of branches. By this I do not mean homoplasy, which is well recognized as an expected result of macro-evolution (Mickey, 1982; Donoghue, 1990), but support for the immediate two alternative branch arrangements for any internode. These two arrangements are obtained through simply switching one of the branches terminating the internode (of the optimal branch) with the free (not rooted) lineage at its base, then doing the same with the other terminal lineage and the basal free lineage, and is called “nearest neighbor interchange”. Three alternative arrangements—one optimal and two suboptimal—are obtained with any four lineages or taxa in unrooted trees, and the arrangement with the best supported internode represents a parsimonious choice between these three arrangements.

If the optimal (shortest tree) branch arrangement is noted as ((AB)C), the support in numbers of steps for the two alternatives ((AC)B) and ((BC)A) may be determined by recalculation with constraint trees. Comparing the relative support (in steps of advanced character states) for ((AB)C, ((AC)B, and ((BC)A allows evaluation of how much greater is the support in numbers of steps for the optimal branch arrangement over that for the two immediate alternatives, and is the basis for the Conditional Probability of Reconstruction measure (Zander, 2001a).

For instance, a dataset for 20 characters of 22 species of the moss genus *Didymodon* (Pottiaceae) with one species, *Barbula unguiculata*, as outgroup, was presented and cladistically analyzed by Zander (1998), and evaluated with constraint trees (Zander, 2001a) to obtain support for the two alternative arrangements for each internode of the optimum tree. With the null hypothesis that the

number of steps (unweighted) of the sum of the internodes of the entire optimal tree are 1/3 of the sum of the steps supporting all three alternative branch arrangements at each internode, and all resolution is random, one might expect that any significant variation is due to shared ancestry. This null hypothesis may be applied to one internode, any lineage, or the whole cladogram.

In fact, the number of steps of the entire optimal tree, 34 (not including those of the basal internode), is significantly greater than the sum of the steps supporting the pairs of alternative branches, 16. An exact binomial calculation (Lowry, 2002) indicated that the possibility that this ratio (34:16) or better would occur by chance alone (at probability 1/3) is less than one out of a thousand. This yields a confidence level of 99.9+% that considerable phylogenetic signal is present in the optimal tree (as alternative hypothesis).

This same assurance—that the *Didymodon* cladogram is solidly based on shared ancestry—is also obtained with a parametric bootstrap (Huelsenbeck & Hillis, 1996) using randomized datasets. Fifty datasets were generated with a random dataset generator (Zander, 2001b), but otherwise identical to the original dataset (i.e., identical numbers of taxa, of columns, and of states in each column). Parsimony analysis with PAUP* (Swofford, 1998) found that all random datasets produced fully resolved trees (demonstrating that resolution alone gives no support of reliability) that were much longer than the length of the optimal tree. The trees averaged 126 steps in a binomial distribution of the 50 trees between extremes of 118 and 134 steps. The original dataset produced a tree only 63 steps in length. Thus, parametric bootstrapping using random data also demonstrates considerable phylogenetic signal for the tree as a whole.

But what is the reliability of the branch arrangement at individual internodes in the optimal tree? As is usual with morphological data, the data were too few to allow nonparametric bootstrapping. Analysis of the individual internodes showed that there were usually too few data to allow confidence at the 95% level using an exact binomial calculation comparing steps supporting the three arrangements of each internode. The minimum distribution of the relative numbers of synapomorphies support-

ing the three possible branch arrangements at each internode to attain a standard (at least in psychology, Games & Klare, 1967) 95% confidence level (that the arrangement was not due to chance alone) is either a minimum of 3:0 (at least three steps supporting the optimal configuration and none supporting either of the two contrary arrangements), or 4:1, 5:2, 6:3, 7:4, 8:5. Each of these minimum ratios of optimal steps versus summed contrary steps would occur only ca. 4% of the time (Lowry, 2002), yielding ca. 96% confidence level (higher ratios yield slightly higher confidence levels). Only twice in the cladogram was one of these reached, support at 3:0 of the arrangement (((*Didymodon australasiae*, *D. umbrosus*, *D. revolutus*)(*D. brachyphyllus*, *D. nevadensis*)) *D. sinuosus*), and support at 4:1 for ((*D. perobtusus*, *D. subandreaoides*)*D. nigrescens*).

The assertion of that previous paper (Zander, 2001a) that support of 1:0 is reliable is here rejected. In the shortest tree, 12 out of 22 (or 55%) of the internal branches show contradiction (that is, are not 1:0 or 2:0 or 3:0, etc.) and therefore require an explanation of convergence among at least two of the three nearest neighbor lineages. Because more than half of the internal branches of the cladogram demonstrate convergence between closely related lines, it is clear that convergence is also a quite acceptable alternative explanation for at least some of the synapomorphies supporting the eight internal branches that have no support for immediate contradiction. Thus, the potential of convergence between nearest neighbors makes an assumption of reliably shared ancestry dubious for individual internodes with no support for contrary arrangements, even though the extent of nesting shown above by exact binomial calculation for the entire cladogram and by parametric bootstrapping with randomized data sets implies considerable phylogenetic signal for the cladogram as a whole.

The parametric bootstrap demonstrated that the cladogram was much more strongly nested than by chance alone, this being evidence for phylogenetic signal. The whole cladogram may be considered a loaded side of a (three-sided) coin. Any random dataset will soon produce a false rejection of the null when maximum support of any one internode is arbitrarily called the loaded side of the die and these support values in numbers of steps are added across internodes. But, since the cladogram was demonstrated by the parametric bootstrap to be well nested, the support at each internode is part of the evidence of loading of the entire cladogram, and the object is to find the least number of internodes with sufficient evidence to reliably support such loading (i.e., the number of contiguous internodes needed to ensure that there is included at least one arrangement of branches that truly reflects shared ancestry at a 95% confidence level). This can be done by adding the evidence for load-

ing, i.e., steps supporting optimal arrangements, and asking: "What is the chance that this support would occur by chance alone at 1/3 probability compared to the sum of steps for all three arrangements at each internode?" Support at each internode is part of the evidence that the cladogram is loaded (i.e., that the cladogram is, say, "heads"). Just because a single internode favors to some slight extent one optimal arrangement does not mean that, probabilistically, that arrangement is sufficiently more loaded than the other two so that it must have occurred.

The required minimum ratios of support for the optimum arrangement versus the summed pairs of immediate alternative arrangements change from 3:0, to 4:1, to 5:2, to 6:3, to 7:4, to 8:5, etc., for needed significance at the 96% confidence level. There are 22 internodes. The ratio of steps supporting the optimal tree versus those supporting contrary trees through nearest neighbor interchange is 34:16 (or 2.124:1 or 4.25:2 or 6.375:3 or 8.5:4). This best approximates 6:3, and if each internode were one step in length, it would take six contiguous internodes to support a reliable phylogenetic analysis because this would involve only three contrary steps. But there are actually 34 steps and only 22 internodes (not including the basal internode), yielding 1.5 steps per internode as average patristic distance. Thus, four contiguous internodes are required on the average for this cladogram to demonstrate reliable phylogenetic signal.

If a patristic distance involving four contiguous internodes is the phylogenetic resolution of this cladogram, it is entirely possible that other morphologically based cladograms with likewise about as many characters as terminal taxa have a similar resolution. Given that we can now judge the reliability of morphologically based cladograms without resorting to philosophy (simplicity, converging on the truth), how might morphologically based cladograms compare with molecularly based cladograms as to their relative reliability?

Parsimony analysis of large molecular datasets commonly allows nonparametric bootstrapping as a measure of reliability of individual internodes. The level of nonparametric bootstrapping needed for reliability was studied (Zander, in prep.) through comparison of various levels of support for three branch arrangements using artificial 4-taxon datasets with PAUP* (Swofford 1998) with 2000 bootstrap replications. Exact binomial calculations at 1/3 probability were compared to bootstrap proportions. To unambiguously attain 95% confidence level, it was found that bootstrap proportions of at least 95% are needed with an optimum branch length of 5, 91% with a length of 15, 87% with a length of 20–40, and 87% with a length of 45 through 60. It is doubtful, then, that most optimal branch arrangements in the literature with less than 89% bootstrap proportion are reliable.

Assuming that a nonparametric bootstrap proportion of 89% usually approximates a standard statistical confidence level of 95%, trees or lineages with 89% or more bootstrap proportions are then reliable gene trees. Lineage sorting, however, is a problem. Genes may exist polymorphically in populations and their evolution does not necessarily coincide with that of the species (Doyle, 1992; Hudson, 1992; Moore, 1995; Maddison, 1996; Lyons-Weiler & Milinkovitch, 1997). One may use the rationale that different gene histories may be treated as single characters (Doyle, 1992). With the null hypothesis that different gene histories are equiprobable for each of the three lineage arrangements at each internode and all variation is random, a single gene tree requires three contiguous internodes before one can assume that at least one species history is present between the separated taxa at a 95% confidence level, and three contiguous internodes is thus the cladogram's average resolution. If half of the internodes have less than 89% bootstrap proportion, then the phylogenetic resolution is a minimum of six contiguous internodes.

With two gene trees agreeing, two internodes are needed on the average; with half of the internodes at less than a 89% nonparametric bootstrap proportion, a 95% confidence level is attained with a minimum of four contiguous internodes. Only with three reliable gene trees agreeing have we acceptable phylogenetic resolution at only one internode; with half the internodes below a 89% bootstrap proportion, the resolution is two contiguous internodes.

In the case of gene trees, one might try to combine molecular datasets and recalculate. The problem is that random data do generate resolved trees, and better resolution is not indicative of a better phylogenetic reconstruction. Combining data sets when their shortest trees differ potentially mixes different gene histories, and may garble the reconstruction. Also, even if the nonparametric bootstrap proportion increases when data sets of trees that agree are combined, in no case can only two agreeing gene trees be reliable as representing the species tree at a 95% confidence level at the one internode resolution. They simply agree in topology but may not reflect species evolution.

One might also use each tree as a Bayesian prior for succeeding trees. With a 95% credibility interval (judged from nonparametric bootstrapping as detailed above) in each of two agreeing trees, the posterior probability of the gene trees is calculable as 96.55%. With three agreeing gene trees, using the first two as priors yields a final posterior probability of 99.7%. This is, however, not a measure of reliability in phylogenetic reconstruction of the species tree, since we are combining in the same calculation the probabilities that these cladograms correctly reflect their gene tree, not the species tree, which might

be different. There is, with three agreeing gene trees, a 5% chance that those agreeing gene trees have a topology different from that of the optimal tree. The only acceptable measure of the reliability of phylogenetic reconstruction is the chance of these three demonstrably reliable gene trees agreeing not being due to chance alone is a 96% confidence level (being the chance of three or more gene trees agreeing at probability 1/3).

In summary, given the discussion and assumptions above, (1) there are doubtless cases in which morphologically based cladograms are more reliable than molecularly based cladograms, and (2) cladograms presently in the literature reliably demonstrate shared ancestry in large part only at a considerable patristic distance.

LITERATURE CITED

- Avice, J. C.** 1994. *Molecular Markers, Natural History and Evolution*. Chapman and Hall, New York.
- Donoghue, M. J.** 1990. Why parsimony? *Evolution* 44: 1121–1123.
- Doyle, J. J.** 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. *Syst. Bot.* 17: 144–163.
- Games, P. A. & Klare, G. R.** 1967. *Elementary Statistics: Data Analysis for the Behavioral Sciences*. McGraw-Hill, New York.
- Hudson, R. R.** 1992. Gene trees, species trees and the segregation of ancestral alleles. *Genetica* 131: 509–512.
- Huelsenbeck, J. P. & Hillis, D. M.** 1996. Parametric bootstrapping in molecular phylogenetics: applications and performance. Pp. 19–45 in: Ferraris, J. D. & Palumbi, S. R. (eds.), *Molecular Zoology: Advances, Strategies, and Protocols*. Wiley-Liss, New York.
- Kendall, M., Ord, J. K., Stuart, A. & O'Hagan, A.** 1987. *Kendall's Advanced Theory of Statistics*. Edward Arnold, London.
- Lowry, R.** 2002. *VassarStats: Web Site for Statistical Computation*. Department of Psychology, Vassar College, Poughkeepsie, New York. 7 March, 2002. <http://faculty.vassar.edu/lowry/VassarStats.html>
- Lyons-Weiler, J. & Milinkovitch, M. C.** 1997. A phylogenetic approach to the problem of differential lineage sorting. *Mol. Biol. Evol.* 14: 968–975.
- Maddison, W. P.** 1996. Molecular approaches and the growth of phylogenetic theory. Pp. 47–63 in: Ferraris, J. D. & Palumbi, S. R. (eds.), *Molecular Zoology: Advances, Strategies, and Protocols*. Wiley-Liss: New York.
- Mickevich, M. F.** 1982. Transformation series analysis. *Syst. Zool.* 31: 461–478.
- Moore, W. S.** 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution* 49: 718–726.
- Philippe, H., Lecointre, G., Hoc Lanh Vân Lê & Le Guyader, H.** 1996. A critical study of homoplasy in molecular data with the use of a morphologically based cladogram, and its consequences for character weighting. *Mol. Biol. Evol.* 13: 1174–1186.

-
- Seberg, O., Petersen, G. & Baden, C.** 1997. Taxonomic incongruence—a case in point from plants. *Cladistics* 13: 180.
- Sites, J. W., Davis, Jr., S. K., Guerra, T., Iverson, J. B. & Snell, H. L.** 1996. Character congruence and phylogenetic signal in molecular and morphological data sets: a case study in the living iguanas (Squamata, Iguanidae). *Mol. Biol. Evol.* 13: 1087–1105.
- Swofford, D. L.** 1998. *PAUP**. *Phylogenetic Analysis Using Parsimony (* and Other Methods)*. Ver. 4. Sinauer Associates, Sunderland, Massachusetts.
- Zander, R. H.** 1998. A phylogrammatic evolutionary analysis of the moss genus *Didymodon* in North America north of Mexico. *Bull. Buffalo Soc. Nat. Sci.* 36: 81–115.
- Zander, R. H.** 2001a. A conditional probability of reconstruction measure for internal cladogram branches. *Syst. Biol.* 50: 425–437.
- Zander, R. H.** 2001b. *Randset: Random Data Set Generator*. April 20, 2001. Buffalo Museum of Science. http://www.buffalomuseumofscience.org/div_botany.htm, also available from the author.