

Imperfect Information and the Balance of Cladograms and Phenograms

STEPHEN B. HEARD¹ AND ARNE Ø. MOOERS

Department of Zoology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

Do cladograms tend to be more imbalanced (less symmetrical) than phenograms? Colless (1982) suggested anecdotally that they do, but more rigorous examination of 208 literature cladograms and phenograms (Heard, 1992; see also Mooers, 1995) revealed no such difference. In a recent note, Colless (1995) returned to this question and concluded that when cla-

distic and phenetic techniques are used to estimate phylogenies from the same data sets, the cladistic trees are consistently and strikingly more imbalanced than the phenetic trees. Colless (1995:105) explained this difference by claiming that "PAUP (and by extension the phylogenetic methods [parsimony] it seeks to model) is biased towards producing asymmetrical [imbalanced] dendrograms." We argue instead that Colless's results only show differences in the way cladistic and phenetic techniques handle data that are too sparse

¹ Present address: Department of Biological Sciences, University of Iowa, Iowa City, Iowa 52242, USA
E-mail: stephen-heard@uiowa.edu

or too homoplastic to be informative and that these results do not contradict earlier findings of agreement for more robust trees. Therefore, Colless's results are important for our understanding of errors in phylogenetic reconstruction but should not weaken studies of tree balance or affect the choice of systematic methodology when data are informative.

Colless (1995) used three different kinds of data to examine differences between cladograms and phenograms: empirical data, random data, and computer-generated data. He used cladistic and phenetic techniques to estimate phylogeny for 20 real data matrices, he applied PAUP (Swofford, 1992; a cladistic method) and a variety of phenetic methods to matrices of random data, and he used both techniques to estimate trees for artificially evolved data for which the "true" (computer simulated) phylogeny was known. Here, we reexamine each kind of data.

The random data matrices were constructed for sets of 11 and 21 taxa with 20 and 30 characters, respectively. Each character value was determined at random, and PAUP and several phenetic methods were used to construct trees. Any hierarchical structure in these random matrices had arisen by chance (see Goloboff, 1991, for a discussion) and so there was no question of finding the "real" tree. Under these conditions, Colless's cladistic trees were much more imbalanced than his phenetic ones, and his interpretation was that cladistic methods produce overly imbalanced trees. However, we find fault with the phenetic estimates, not the cladistic ones.

With completely random character data, we have no reason to think any one possible tree more likely than any other. When trees are constructed with parsimony from random data, this is exactly the outcome: all possible trees are equally likely (Slowinski, 1990; Guyer and Slowinski, 1993, J. Slowinski, pers. comm.). The expected distribution of tree topologies for this result is known (the "all labelled trees equiprobable" or "proportional to distinguishable arrangements" null model; Rosen, 1978; Rogers, 1993). Completely random trees

TABLE 1. *I* values (Colless's index of imbalance) for random matrices using five tree construction techniques (Colless, 1995: table 3) and expected *I* for completely random trees (Rogers, 1993). *I* is computed as (Heard, 1992) $I = \sum (\text{all interior nodes}) |T_R - T_L| / ((n-1)(n-2)/2)$, where at each interior node of a tree of size *n* the right and left branches subtend *T_R* and *T_L* tips, respectively.

| Technique ^a | 11 taxa/tree | | 21 taxa/tree | |
|------------------------|--------------|------|--------------|------|
| | \bar{x} | SE | \bar{x} | SE |
| Random expectation | 0.56 | 0.02 | 0.46 | 0.01 |
| Random data | | | | |
| PAUP | 0.63 | 0.11 | 0.49 | 0.06 |
| MNN | 0.40 | 0.07 | 0.69 | 0.05 |
| MFN | 0.18 | 0.04 | 0.11 | 0.01 |
| MUP | 0.31 | 0.04 | 0.23 | 0.03 |
| MWP | 0.31 | 0.03 | 0.17 | 0.02 |

^a PAUP is a cladistic technique. Nearest-neighbor joining (MNN), furthest-neighbor joining (MFN), UPGMA (MUP), and weighted PGMA (MWP) are phenetic techniques using Manhattan distances.

are more imbalanced than expected under common null models of macroevolution (Slowinski, 1990; Rogers, 1993) and more imbalanced than samples of trees from the literature (Guyer and Slowinski, 1993; Cunningham, in press).

Colless's cladistic trees, in fact, match the random expectation very closely in balance (Table 1). In contrast, in seven of eight cases (two tree sizes, four methods) the phenetic techniques produced sets of trees that are significantly more balanced than the random expectation. Clearly the phenetic trees cannot be better reconstructions. Instead, when data are random, cladistic methods (appropriately) produce trees matching the random expectation. Phenetic methods, however, sample from a more balanced (and incorrect) population of possible trees (see also Huelsenbeck and Kirkpatrick, 1996). This difference in how the techniques deal with poor data also explains Colless's other results.

Colless's (1995) second kind of data was a set of 20 real data matrices, to which he applied a battery of phylogeny-estimation techniques (because in several cases two or more matrices consisted of different sets of characters from the same taxa, the 20 matrices represented only 11 independent phylogenies). Again, the apparent result is

that the cladistic trees are strikingly more imbalanced than the phenetic trees. However, there are two major reasons for concern.

First, the 11 phylogenies are neither a random nor a systematic sample from the population of published phylogenies (in contrast to Heard's [1992] and Mooers's [1995] trees), and it is therefore impossible to generalize beyond the handful of data matrices presented. Second (even if we disregard the first point), most of the tree estimates based on these matrices are not robust because the data are either sparse or extremely homoplastic. Colless (1995) did not report any direct measures of robustness, but most of the 20 trees appear to be very poorly supported. Of 11 trees for which we could find some indication of data quality, only two (BEG, BIL) are likely to be at all well supported and nine (H1A, H1B, H2A, H2B, P1A, P1B, P2A, P2B, ALL) must be very poor estimates of true phylogeny because among trees for the same clades based on different suites of characters there is little or no consensus (Rohlf et al., 1983: fig. 1, table 3). The remaining nine trees are unlikely to be much better; their data matrices have on average just 1.4 informative characters/taxon (D. Colless, pers. comm.).

The issue of tree robustness is important because Huelsenbeck and Kirkpatrick (1996) have shown that when phylogenetic information is very good, both phenetic and cladistic techniques produce trees with very close to "correct" imbalance. However, as phylogenetic information becomes poorer, cladistic topologies move towards the expectation for completely random data, which ordinarily means becoming more imbalanced (Mooers et al., 1995; Huelsenbeck and Kirkpatrick, 1996). Colless (pers. comm.) has also run simulations in which cladistic topologies for extremely imbalanced true trees moved toward the more balanced random expectation. As a group, Colless's 20 cladistic trees are not distinguishable from the random expectation (sign test, $P = 0.5$ if we ignore the nonindependence issue). UPGMA phenetic trees, however, do not

move toward random topologies as data deteriorate. In a nutshell, when data are poor, cladistic estimates look like the random guesses they should be; phenetic estimates are no better at getting a correct tree (Huelsenbeck and Kirkpatrick, 1996) and they err toward balanced estimates. The pattern in Colless's real data matrices is exactly the pattern seen previously in the entirely random matrices.

The third kind of data mustered by Colless (1995) was a set of artificially evolved trees, for which both topology and matrices of character data were obtained under reasonable models of evolution. Trees estimated by cladistic and phenetic methods could then be compared with each other and with the "real" trees by their imbalance values (I), and again cladistic trees appear consistently more imbalanced than phenetic ones. Does this mean that the cladograms are too imbalanced? On closer examination, the answer is no. Colless's artificially evolved trees included 20 species and were estimated from data matrices of 20 characters; these are very sparse data. Colless broke his trees down into a set of 10 imbalanced trees ($0.49 \leq I \leq 1.0$ for the "real" trees) and 10 balanced ones ($0.13 \leq I \leq 0.29$), but the cladistic estimates do not differ in balance between the two groups ($t_{18} = 1.35$, $P = 0.19$) and as an aggregate do not differ from the random expectation ($I = 0.48 \pm 0.03$ [$\bar{x} \pm \text{SE}$] vs. expected [for 20 taxa] $I = 0.47 \pm 0.02$). Similar results are obtained from data sets with more characters but with severely homoplastic data (D. Colless, pers. comm.). In other words, estimated trees based on very poor data look little better than random guesses. Like his real data matrices, Colless's artificially evolved trees tell the same story as his random data matrices. Colless mentioned data from a third set of trees generated with five times as many characters; for these trees, both cladistic and phenetic estimates had balance "close to or identical with" the real value. Colless called this "a very strange result," but we do not find it strange at all—it lines up well with other results showing agreement among methods for more reliable trees (Heard, 1992;

Mooers et al., 1995; Huelsenbeck and Kirkpatrick, 1996).

In summary, we do not believe Colless's (1995) data show that cladistic and phenetic trees always differ in imbalance, and we do not believe that cladistic methods perform improperly with respect to imbalance. Rather, Colless's data show that when data are poor and therefore estimates are unreliable, the errors made by cladistic and phenetic methods differ. When data are so poor that the true phylogeny cannot be recovered, cladistic methods offer estimates that look like random guesses among all possible trees. Phenetic methods, however, preferentially produce estimates that are more balanced although not more correct. We agree with Colless that there is a difference between phenetic and cladistic methods but disagree that he has shown that this difference applies to robust trees. The difference does serve to underline the importance of considering the robustness of trees when examining patterns in balance.

What about differences between cladograms and phenograms when both are well supported? Such a difference would be important because it would indicate that at least one technique is failing to provide useful estimates of real trees. We expect that when trees are robust, the methods will tend to converge on the correct topology and therefore not differ much in balance (e.g., Huelsenbeck and Kirkpatrick, 1996). We do acknowledge that it has not yet been established whether, in practice, the methods converge entirely or merely approximately or whether they converge from the same or different ends of the balance spectrum. The relative balance of cladograms and phenograms, when these are reasonably reliable attempts at tree estimation, remains open to question, but we currently see little evidence that it differs much.

ACKNOWLEDGMENTS

We thank M. MacInnis, D. Schluter, and J. Slowinski for comments. J. S. Rogers supplied expected *I* values for the "proportional to distinguishable arrangements" null hypothesis. We are particularly grateful to D. Colless for discussion and for sharing results of several new simulations. R. Yorke pointed out (with considerable relish) our overuse of "mustered." We were supported by Izaak Walton Killam Postdoctoral Fellowships and by E. B. Eastburn Postdoctoral Fellowships from the Hamilton Foundation.

REFERENCES

- COLLESS, D. H. 1982. Review of *Phylogenetics: The Theory and Practice of Phylogenetic Systematics*, by E. O. Wiley. Syst. Zool. 31:100-104.
- COLLESS, D. H. 1995. Relative symmetry of cladograms and phenograms: An experimental study. Syst. Biol. 44:102-108.
- CUNNINGHAM, S. A. In press. Problems with null models in the study of phylogenetic radiation. Evolution.
- GOLOBOFF, P. A. 1991. Random data, homoplasy, and information. Cladistics 7:395-406.
- GUYER, C., AND J. B. SLOWINSKI. 1993. Adaptive radiation and the topology of large phylogenies. Evolution 47:253-263.
- HEARD, S. B. 1992. Patterns in tree balance among cladistic, phenetic, and randomly generated phylogenetic trees. Evolution 46:1818-1826.
- HUELSENBECK, J. P., AND M. KIRKPATRICK. 1996. Do phylogenetic methods produce trees with biased shapes? Syst. Biol. 45 (in press).
- MOOERS, A. Ø. 1995. Tree balance and tree completeness. Evolution 49:379-384.
- MOOERS, A. Ø., R. D. M. PAGE, A. PURVIS, AND P. H. HARVEY. 1995. Phylogenetic noise leads to unbalanced cladistic tree reconstructions. Syst. Biol. 44:333-343.
- ROGERS, J. S. 1993. Response of Colless's tree imbalance to number of terminal taxa. Syst. Biol. 42:102-105.
- ROHLF, F. J., D. H. COLLESS, AND G. HARI. 1983. Taxonomic congruence re-examined. Syst. Zool. 32:144-158.
- ROSEN, D. E. 1978. Vicariant patterns and historical explanation in biogeography. Syst. Zool. 27:159-188.
- SLOWINSKI, J. B. 1990. Probabilities of *n*-trees under two models: A demonstration that asymmetrical interior nodes are not improbable. Syst. Zool. 39:89-94.
- SWOFFORD, D. L. 1992. PAUP: Phylogenetic analysis using parsimony, version 3.0S+1. Illinois Natural History Survey, Champaign.

Received 7 June 1995; accepted 12 August 1995
Associate Editor: David Cannatella