Systematic Botany (1996), 21(2): pp. 237–253

Commentaries

Accuracy of Parsimony Analysis Using Morphological Data: A Reappraisal.—ABSTRACT. In a recent paper, W. F. Lamboy used computer simulations to evaluate the accuracy of parsimony analysis using morphological characters. Based on his simulations, Lamboy concluded that parsimony analysis of morphological data is generally highly inaccurate, that many published trees are likely incorrect, and that current emphasis on phylogeny reconstruction using parsimony is misplaced. In this paper we argue that these conclusions are the result of: 1) using an unusually small number of characters relative to the number of taxa; 2) generating true trees that contain polytomies; 3) using misleading measures of accuracy; 4) summing across a biased set of simulation results to make general conclusions; and 5) using an inappropriate index to compare simulated data to results in the real world. We provide recommendations for designing more rigorous simulations of phylogenetic questions, and summarize current knowledge of the performance of parsimony analysis based on other simulation studies. In contrast to Lamboy, we see great promise for accurate phylogeny reconstruction using parsimony analysis of morphological data.

Systematists have many difficult choices to make when undertaking a phylogenetic analysis. For example, they must decide what characters to sample, what species to include, and what method they will use to build trees. Computer simulations offer an important means for testing how different approaches affect the accuracy of phylogeny reconstruction. Simulation studies allow one to specify a known phylogeny and to control the conditions under which the phylogeny and characters are generated (see Huelsenbeck 1995). Thus, one can determine how variation in the parameters of interest affects the performance of a method. Simulations have been used to address a variety of systematic questions, including the relative success of different reconstruction methods (see Hillis 1995, for a recent summary), the validity of methods for testing the accuracy and signal of trees and data sets (Hillis and Huelsenbeck 1992; Zharkikh and Li 1992a,b; Hillis and Bull 1993; Harshmann 1994), the benefits of including fossil versus living taxa (Huelsenbeck 1991b), the consequences of combined versus separate analysis of data sets (Bull et al. 1993a), and the effects of concerted evolution on phylogenetic analysis using multiple genes (Sanderson and Doyle 1992). Most recent studies have simulated DNA sequence data and older studies have simulated allele frequency data (Nei et al. 1983; Rohlf and Wooten 1988; Kim and Burgman 1988). Few simulation studies have explicitly simulated morphological data (Lynch 1989), although some studies have simulated generalized two-state or multistate characters that may be interpreted as morphological data (Rohlf et al. 1990; Hillis and Huelsenbeck 1992). This paucity of explicit simulations of morphological characters is unfortunate, because morphology remains the most common type of character data used in phylogenetic analyses (Sanderson et al. 1993).

This deficiency was recently addressed by Lamboy (1994), who undertook a simulation study to evaluate the accuracy of parsimony analysis using morphological data. Lamboy concluded that parsimony analysis of morphology was often highly inaccurate, and extrapolated that many published phylogenetic studies are probably incorrect. In this paper, we argue that Lamboy's study gives a highly misleading picture of morphology, the performance of parsimony, and the current "state of the art" in computer simulations of phylogenetic questions. In contrast to Lamboy, we see great promise for accurate phylogeny reconstruction using parsimony analysis of morphological data, and we give a number of recommendations for how to evaluate and improve its success.

What is Special About Morphology?

Most recent simulation studies have simulated DNA data as opposed to morphology. Given this, one might reasonably ask: what special properties of morphological data have been ignored that will affect their success in reconstructing the true phylogeny? Lamboy (1994: 490) emphasized several differences, namely: 1) relatively few character states per character; 2) few characters relative to the number of species, and 3) "reversals and parallelisms." None of these properties are unique to morphological data. In his analyses, Lamboy used either two states per character (which occurs variably in both morphological and DNA studies), or four unordered states (the maximum number typically observed in DNA characters). Furthermore, both morphological and molecular characters can potentially have a large number of character states, such as a complex and variable morphological feature or an allozyme locus with a large number of alleles. The ratio of characters to taxa is highly variable in molecular and morphological studies, but the relationship between number of characters and accuracy (for a given number of taxa) is clear. Except under certain combinations of branch lengths (Felsenstein 1978), increasing the number of characters generally increases accuracy (Nei et al. 1983; Kim and Burgman 1988; Rohlf and Wooten 1988; Charleston et al. 1994; Hillis et al. 1994a,b). In half of Lamboy's simulations, the ratio of characters to taxa is 1:1 (if uninformative charaters are included; otherwise, there are more taxa than characters). This relatively small number of characters clearly predetermines the poor performance of parsimony, but is found only rarely in real data sets—only 7.5% of the morphological data sets surveyed by Sanderson and Donoghue (1989) have so few informative characters, and the mean character to taxon ratio is 2.9 characters per taxon for the 40 real data sets they surveyed. Reversals and parallelisms occur in both molecular and morphological data, and have consistently been included in previous simulation studies. In sum-

mary, the data sets generated by Lamboy have no special properties that make them more relevant to morphological studies than studies that have simulated DNA, allele frequencies, or generalized character data.

Are there any general and unique properties of morphological data that could affect accuracy and that could have been simulated? One difference is that the heritability of morphological characters is usually unknown, allowing the possibility that some characters do not have a genetic basis and thus could be misleading about phylogeny. Furthermore, many morphological characters involve continuous variation (whether they are recognized as such or not; Stevens 1991). One could also argue that molecular frequency data are continuous, but discrete morphological characters can be treated as frequencies as well (Wiens 1993, 1995). Thus, the presence of continuous variation (such as the length of a stamen or bone) in morphology appears to be another legitimate difference. Although not cited by Lamboy, the accuracy of phylogenetic analysis using continuous morphological variation has been addressed explicitly in a previous simulation study (Lynch 1989).

Other differences between molecular and morphological data are not so clear. For example, non-independence has been postulated for both molecular (Wheeler and Honeycutt 1988; Korber et al. 1993) and morphological data (Winterbottom 1990), and is a general problem that needs to be addressed in more simulation studies. Fossils offer special problems and rewards, but are no longer restricted to morphology (DeSalle et al. 1992), and have received some attention in simulation studies (Huelsenbeck 1991b). Although most molecular characters do not have ontogenies, the legitimate role of ontogenetic information in phylogenetic analysis, aside from being a source of new characters, has been greatly overestimated (Mabee 1993).

Instead of exploring any of these properties, Lamboy chose to vary primarily a parameter (number of characters) which has a relatively well-known effect on accuracy. The effects of other parameters he varied (number of states per character, polytomous speciation, different types of homoplasy) is not always obvious from other studies, but Lamboy's assessment of their impact on accuracy is seriously compromised by the design of his study.

Problems of Design: To Estimate the Impossible Tree

In Lamboy's simulations, many of the true trees contained polytomies. Lamboy stated (p. 491) that these unresolved trees "provide a more realistic array of tree topologies than the sets of completely resolved trees used in many previous studies." We seriously question the idea that the simultaneous splitting of eight species (p. 491) is in any way "realistic." Yet, Lamboy's description of his tree generating method (p. 491) suggests that polytomous splitting occurred frequently in his simulated phylogenies. Furthermore, Lamboy considered a methods ability to leave these polytomies unresolved to be an important component of accuracy. This is certainly a dramatic departure from previous notions of the concept (see review by Hillis 1995). Most phylogenetic approaches assume that the tree is dichotomous, and resolve the tree wherever support exists (there are parsimony methods that do not resolve polytomies if there is insufficient signal to do so, but these methods were not examined by Lamboy; Crandall et al. 1994). Although it might have been interesting to see how violations of the dichotomy assumption affect accuracy, this information is unavailable from Lamboy's paper. Instead, this basic assumption is violated throughout the simulations with unknown (but seemingly high) frequency. Thus, the problem of polytomous model trees represents a confounding factor that by itself would seriously compromise the validity of the conclusions.

Measuring Accuracy: All or Nothing

Lamboy used several measures to assess the accuracy of the estimated trees, including "percent of phylogenies for which the true tree was among the set of most parsimonious," "percent of most parsimonious trees that were the true tree," and "percent of strict consensus trees that were the true tree." The common theme among these measures is that an estimated tree is either identical to the true tree (including any polytomies) or is wrong. Although this "all or nothing" approach to measuring accuracy is appropriate for trees with a single internal stem (i.e., a four taxon unrooted tree; e.g., Kim and Burgman 1988; Huelsenbeck and Hillis 1993), it is not appropriate for a tree with eight species, as in Lamboy's simulations. For eight taxa, the estimated tree can be wrong but still have an undeniable similarity to the correct tree, or could be wrong and have no nodes in common with the true tree (Fig. 1). Lamboy's criteria would give these incorrect trees equal weight. The dismal picture of the success of parsimony and morphology that Lamboy paints is due in large part to these insensitive and misleading measures of accuracy. Hillis et al. (1994a) identified this problem as one of the principal sources of bias in simulation studies; to avoid this bias, they recommended that methods be scored using the average number of correctly resolved components among all equally optimal solutions.

Lamboy's use of polytomous trees as true phylogenies confounded the measuring of accuracy in other ways as well. Lamboy considered the "percent of time the strict consensus tree was identical to the true tree as the primary measure of accuracy" (p. 489). Strict consensus trees are rarely interpreted as estimates of phylogeny, probably because they are simply collapsing nodes of disagreement. Using this index of accuracy, all the most parsimonious trees might contradict the true tree, but the resulting consensus tree would be considered "correct." Thus, using the consensus tree as a measure of accuracy confuses polytomies due to simultaneous speciation with those due to character conflict.

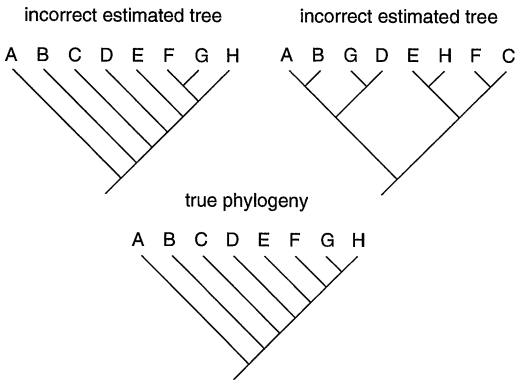

FIG. 1. An example showing that two estimated trees can be incorrect but nevertheless differ greatly in their similarity to the true phylogeny. This illustrates the need to use a measure of accuracy that takes into account the proportion of correctly resolved clades, not simply whether a tree is right or wrong.

PROBLEMS OF INTERPRETATION

The main conclusion of Lamboy's study was that parsimony rarely finds the correct tree for morphological data sets (p. 489). This conclusion was based on summing his measure of accuracy across all of his simulations. This unorthodox practice carries with it the implicit assumption that the frequency of the various conditions in all the simulations is somehow meaningful, or at least unbiased. Yet, in half the simulations the ratio of informative characters to taxa was unusually low (one informative character per species or fewer). Prior simulation studies have shown the relationship of number of characters to accuracy time and time again, a conclusion also noted by Lamboy. Clearly, summing across the simulations gives neither a realistic nor unbiased picture of the performance of parsimony; values will shift markedly one way or the other depending on the sample of conditions simulated.

After reporting the summed results of his simulations, Lamboy used his results to infer how often real morphological data sets estimate the true tree using parsimony. Rather than reporting how often simulated data sets with comparable numbers of characters estimated the true phylogeny, Lamboy used the consistency index of real and simulated data sets to guide his comparisons. This was an extremely poor choice of indices, because the consistency index is highly sensitive to the number of taxa in an analysis (Archie 1989a,b; Sanderson and Donoghue 1989, and subsequent authors). This property of the consistency index is uncontroversial (contra Lamboy, p. 493), and renders Lamboy's comparisons invalid; for example, he compared his data sets of eight species to real data sets with as many as 68. Lamboy compounded this error by comparing consistency indices of data sets with uninformative characters removed (from the literature) with his data sets, which included uninformative characters. This artificially inflated the measure for his data sets. Thus, Lamboy's claim that "maximum parsimony probably performs more poorly with most real morphological data sets than it did with most simulated data sets analyzed" is completely unsubstantiated.

Finally, Lamboy concluded (p. 502) that "I advise caution in using the results of phylogeny reconstruction to make inferences about any biological processes or patterns that depend upon exact knowledge of the historically true phylogeny." We wholeheartedly agree with this statement. Yet, systematists and users of phylogenies are not so helpless in evaluating the robustness of their results. Methods now exist to statistically distinguish between well and poorly-supported nodes (e.g., bootstrapping, Felsenstein 1985; Faith's 1991 PTP test) and whether or not whole data sets contain useful phylogenetic information (e.g., g_1 index, Hillis 1991; Archie's 1989a randomization test). The performance of many of these methods has now been tested with simulated and/or experimental phylogenies (e.g., g_1 index: Hillis 1991; Huelsenbeck 1991a; Hillis and Huelsenbeck 1992; bootstrapping: Zharkikh and Li 1992a,b; Hillis and Bull 1993; Harshmann 1994). Based on his description of his data-generating methods, we question whether Lamboy's simulated trees actually contain any nodes that are simultaneously wrong and strongly supported. Given that weak support for a clade can be detected and taken into consideration, we believe that it represents a relatively trivial problem in phylogenetic inference. Furthermore, although the users of phylogenetic trees must be aware of the ambiguities in the trees that they use, we note that the use of phylogenies to test evolutionary questions need not require a fully-resolved or strongly supported tree (DeBry 1992; Purvis and Garland 1993; Losos 1994).

BUILDING BETTER SIMULATIONS

Given the many problems of Lamboy's study, can we make practical suggestions as to how to design a better computer simulation study? We can, and we believe that more rigorous simulation studies have already become standard for the field (e.g., Kim and Burgman 1988; Huelsenbeck and Hillis 1993). There are at least four important ingredients to a good simulation study: 1) use of a specific (and specified) model (or models) of evolution to generate the data; 2) unbiased treatment of the parameters potentially important in determining performance; 3) comprehensive examination of the effects of variation in the parameters of interest, and 4) unbiased evaluation of the results.

Examples of different models of evolution that have been used to generate character data for simulations include allele frequencies evolving by random genetic drift (Nei et al. 1983; Kim and Burgman 1988; Rohlf and Wooten 1988) and DNA sequences evolving by the Jukes-Cantor and Kimura mutation models (Huelsenbeck and Hillis 1993; Huelsenbeck 1995). Results of phylogenetic

simulations are known to be highly dependent on the model of evolution (e.g., Rohlf et al. 1990; Wheeler 1992; Huelsenbeck 1995). Because of this fact, simulators need to specify the model of evolution and, ideally, examine the effects of changing the model. Results should not be considered to be general unless they are shown to be relatively insensitive to details of the model. Lamboy did not specify an explicit evolutionary model, and he tried to draw general conclusions from results that appear to be highly specific to the details of his simulations.

Parameters that have already been shown to be important in determining accuracy include: 1) relative branch lengths (Felsenstein 1978); 2) absolute branch lengths (Huelsenbeck and Hillis 1993); 3) number of characters (Kim and Burgman 1988; Charleston et al. 1994; Hillis et al. 1994ab), and 4) tree shape (Fiala and Sokal 1985; Rohlf et al. 1990). These parameters must be set or varied in an unbiased fashion, to avoid predetermining the results from the outset (Huelsenbeck 1995). Furthermore, results concerning the performance of a method must always be reported with the caveat that they may apply only to the specific conditions under which they were generated. For example, Huelsenbeck and Hillis (1993) qualified their results with the warning that they may apply only to the four-taxon case (hence their paper's title). In contrast, Lamboy did not emphasize in his abstract that parsimony analysis of morphological data performs poorly when the number of characters is small, but rather that it generally performs poorly.

Because factors such as tree shape and branch lengths can bias the results of simulation studies, it is important to define the parameter space of interest and analyze it as completely as possible. Biased selection of model trees, for instance, has led to biased conclusions about the relative success of different methods. For example, by analyzing only trees with branch lengths known to positively mislead parsimony, Tateno et al. (1994) mistakenly concluded that distance correction is superior to weighted parsimony at high rates of evolution. This highlights the need for thorough examination of problems and explaining the results of simulations in context. In the case of the relationship of number of characters to phylogenetic success, character number can be varied systematically for a given set of model conditions, so that the minimum number of characters needed to find the correct tree with high probability can be identified (Hillis et al. 1994b). Of course, this finding will be limited to the model examined and the tree evaluated, so even in this case the results should not be overgeneralized.

Finally, the results of a study must be analyzed in as unbiased a manner as possible. We have already identified at least two sources of bias in Lamboy's analysis. First, his measures of phylogenetic accuracy were unrealistic, uninformative, and misleading. Second, his method of comparing his results to real studies of morphology was based on an inappropriate measure that in any case was calculated differently among the relevant studies. These problems combine to make Lamboy's study largely uninterpretable; certainly his conclusions are not justified by his data.

Parsimony and Accuracy: What is Known?

Given that we have objections to Lamboy's characterization of the performance of parsimony, how would we describe (in general) the accuracy of parsimony based on simulation studies? First, parsimony is highly sensitive to certain extreme combinations of branch lengths, where branch length is defined as the amount of actual, estimated, or potential character change (summed across characters) for a given lineage. In particular, parsimony is inconsistent (i.e., adding characters increases support for an incorrect clade) when two or more unrelated terminal branches are very long compared to adjacent terminal or internal branches (see Fig. 2). This property of parsimony was first described by Felsenstein (1978) and has been shown with computer simulations of gene frequency data by Kim and Burgman (1988) and more recently for DNA sequence data by Huelsenbeck and Hillis (1993) and others. Although some other phylogenetic methods (e.g., maximum likelihood) are consistent when there are long, unrelated branches (i.e., adding more characters eventually leads to estimating the correct tree), all methods are highly inefficient under these conditions (they require a very large number of characters to get the true phylogeny). Furthermore, most other methods are biased in the same direction as is parsimony, such that the long branches will tend to appear together in the estimated trees (Hillis et al. 1994b; Huelsenbeck 1995).

Parsimony (and other methods) also performs poorly when there are few variable characters, and when branches become so long or characters are evolving so quickly that the phylogenetic information present is effectively randomized (Hillis and Huelsenbeck 1992; Huelsenbeck and Hillis 1993; Huelsenbeck 1995). These latter two problems are less severe, because poor character support for a clade and data sets that approach random levels of noise can be easily detected. The problem of long-branch attraction is more difficult, because as more data are added, character support (and confidence) for the incorrect node will increase (Zharkikh and Li 1992b; Hillis and Bull 1993). However, the problem potentially can be solved by adding more taxa to subdivide the long branches (Hendy and Penny 1989; Swofford and Olsen 1990), and it is not known whether long-branch attraction is actually common in real data sets.

Although the branch-length problem may be characterized as relatively well-understood, several questions remain, namely that 1) all studies of branch-length problems so far have allowed for only a limited set of possible branch-length combinations; 2) the effects of different numbers of taxa have not been addressed in a critical fashion, and 3) the application of these studies to determining when long branches are attracting in a given data set in the real world is still unclear. However, parametric bootstrapping (Bull et al. 1993b; Hillis et al.

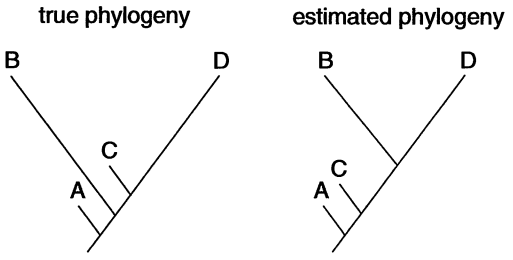

FIG. 2. The problem of long-branch attraction. Parallel changes in taxa B and D are interpreted as support for grouping these two taxa, and the addition of more characters tends to increase support for this incorrect clade. Thus, long branches cause parsimony (and many other phylogenetic methods) to be inconsistent and positively misled.

1994a; Huelsenbeck et al., in press) is beginning to be used to evaluate the performance of phylogenetic methods for particular data sets, and is particularly useful for detecting potential branch-length probems.

Many other problems in parsimony analysis have now been addressed using computer simulations, but many of these remain unresolved. For example, character weighting based on differences in rates of change among characters generally improves parsimony analysis (Huelsenbeck and Hillis 1993; Chippindale and Wiens 1994; Huelsenbeck 1995), although the effects of imperfect knowledge of actual rates (the typical situation with real data) are poorly known, as is the success of many weighting schemes [i.e., dynamically weighted parsimony and combinatorial, EOR (expected to observed ratio), and successive weighting]. Taxa that are relatively incomplete (missing character observations) generally decrease accuracy (Huelsenbeck 1991b), although this decrease can be offset if the taxa are older. Simulations have alternatively shown that adding taxa may increase (Wheeler 1992) or decrease (Charleston et al. 1994) accuracy, even when the number of characters is held constant. Combining data sets may reduce the accuracy of parsimony analysis relative to separate analysis of "good" and "bad" data sets (Bull et al. 1993a), although whether this means it is generally better not to combine data sets or to include more homoplastic data is still questionable (Chippindale and Wiens 1994; Wiens and Chippindale 1994). Concerted evolution in multigene families generally causes errors in parsimony analysis of molecular data at intermediate levels of gene conversion (Sanderson and Doyle 1992). Basic problems that remain to be addressed using computer simulations include the non-independence of lineages and characters and the consequences of different methods for treating intraspecific variation (Wiens 1995).

CONCLUSIONS AND RECOMMENDATIONS

In his study, Lamboy questioned the current emphasis on phylogenetic research because of purportedly finding that parsimony analysis of morphological data performs poorly in simulation. However, this pessimistic view is the result of 1) using an unrealistically small number of characters relative to the number of taxa in half of his simulations, a situation well-known to give inaccurate results; 2) generating true trees that are unrealistic and nearly impossible to reconstruct because they contain numerous polytomies, and requiring that estimated trees include these polytomies to be considered correct; 3) using a misleading and largely uninformative measure of accuracy (counting only whether estimated trees are exactly right or wrong, rather than counting how many nodes are correctly recovered); 4) summing across a biased set of simulations to make general conclusions, and 5) using an inappropriate and inconsistently calculated index to compare simulated data to results in the real world. Furthermore, Lamboy simulated data with no special properties of morphological characters, so previous simulation studies that do not share the deficiencies of Lamboy's study actually provide a much more realistic picture of the success of phylogenetic analyses, morphological or otherwise. These previous studies indicate that the current emphasis on phylogenetic research is supported by methods that perform well under a wide variety of conditions, whether the source of the data is molecular or morphological.

Although we disagree with Lamboy's pessimism, we cannot claim to know whether most reported trees based on parsimony analysis of morphological data are right or not. However, we suspect that many morphological trees are mostly right—how else could one explain the general observation that so many clades strongly supported by morphological data are also strongly supported by molecular data (Donoghue and Sanderson 1992)? Futhermore, the fact that levels of homoplasy are so similar between molecular and morphological data sets (Sanderson and Donoghue 1989) suggests that if trees based on morphology are not generally accurate, they are no more inaccurate than trees based on other kinds of data.

In many ways, molecular systematists are ahead of morphologists in being concerned about accuracy. For example, methods for testing for support and signal were seemingly developed with molecular data in mind, and have been slow in becoming widely used by morphologists. We argue that all systematists need to be aware of basic problems such as: 1) weakly-supported nodes (because of few and/or conflicting characters); 2) data matrices with random levels of noise; 3) long branch attraction; 4) character non-independence, and 5) non-independence of lineages (e.g., lateral gene transfer, hybridization). Statistical methods have been developed that potentially can detect some of these problems, including 1) bootstrapping and Faith's (1991) PTP test; 2) the g1 index and Archie's (1989a) randomization test; 3) parametric bootstrapping, and 4) Maddison's (1990) test of correlated evolution. Other problems, common in or specific to morphological data, also can have a significant impact on results and therefore need to be

considered, such as the problem of defining and ordering morphological character states in an unbiased manner (Stevens 1991; Mabee 1993), the treatment of continuous (Thiele 1993) and discrete intraspecific variation (Wiens 1995), and the use of fossil or incomplete taxa (Donoghue et al. 1989; Huelsenbeck 1991b; Wiens and Reeder 1995). All of these considerations will play a role in making parsimony analysis of morphological data more rigorous and more accurate.

ACKNOWLEDGMENTS. We thank Chris Brochu and the other members of the Systematics Discussion Group at the University of Texas at Austin for useful discussion of Lamboy's paper. Phillip Chu and Brad Livezey provided valuable comments on the manuscript. J. J. Wiens was supported during preparation of the manuscript by a NSF Graduate Fellowship.

LITERATURE CITED

ARCHIE, J. W. 1989a. A randomization test for phylogenetic information in systematic data. Systematic Zoology 38: 239–252.

———. 1989b. Homplasy excess ratios: New indices for measuring homoplasy in phylogenetic systematics and a critique of the consistency index. Systematic Zoology 38: 253–269.

BULL, J. J., J. P. HUELSENBECK, C. W. CUNNINGHAM, D. L. SWOFFORD, and P. J. WADDELL. 1993a. Partitioning and combining data in phylogenetic analysis. Systematic Biology 42: 384–397.

———, C. W. CUNNINGHAM, I. J. MOLINEUX, M. R. BADGETT, and D. M. HILLIS. 1993b. Experimental molecular evolution of bacteriophage T7. Evolution 47: 993–1007.

CHARLESTON, M. A., M. D. HENDY, and D. PENNY. 1994. The effects of sequence length, tree topology, and number of taxa on the performance of phylogenetic methods. Journal of Computational Biology 1: 133–151.

CHIPPINDALE, P. T. and J. J. WIENS. 1994. Weighting, partitioning, and combining characters in phylogenetic analysis. Systematic Biology 43: 278–287.

CRANDALL, K. A., A. R. TEMPLETON, and C. F. SING. 1994. Intraspecific phylogenetics: Problems and solutions. Pp. 273–297 in Models in phylogeny reconstruction, eds. R. W. Scotland, D. J. Seibert, and D. M. Williams. Oxford: Clarendon Press.

DEBRY, R. W. 1992. Biogeography of New Word taiga-dwelling *Microtus* (Mammalia: Arvicolidae): A hypothesis test that accounts for phylogenetic uncertaintly. Evolution 46: 1347–1357.

DESALLE, R., J. GATESY, W. WHEELER, and D. GRIMALDI. 1992. DNA sequences from a fossil termite in Oligo-Miocene amber and their phylogenetic implications. Science 257: 1933–1936.

DONOGHUE, M. J. and M. J. SANDERSON. 1992. The suitability of molecular and morphological evidence in reconstructing plant phylogeny. Pp. 340–368 in Molecular systematics of plants, eds. P. S. Soltis, D. E. Soltis, and J. J. Doyle. New York: Chapman and Hall.

———, J. A. DOYLE, J. GAUTHIER, A. G. KLUGE, and T. ROWE. 1989. The importance of fossils in phylogeny reconstruction. Annual Review of Ecology and Systematics 20: 431–460.

FAITH, D. P. 1991. Cladistic permutation tests for monopyly and nonmonophyly. Systematic Zoology 40: 366–375.

FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Systematic Zoology 27: 401–410.

———. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39: 783–791.

FIALA, K. L. and R. R. SOKAL. 1985. Factors determining the accuracy of cladogram estimation: Evaluation using computer simulation. Evolution 39: 609–622.

HARSHMANN, J. 1994. The effect of irrelevant characters on bootstrap values. Systematic Biology 43: 419–424.

HENDY, M. D. and D. PENNY. 1989. A framework for the quantiative study of evolutionary trees. Systematic Zoology 38: 297–309.

HILLIS, D. M. 1991. Discriminating between phylogenetic signal and random noise in DNA sequences. Pp. 278–294 in Phylogenetic analysis of DNA sequences, eds. M. M. Miyamoto and J. Cracraft. New York: Oxford Univ. Press.

———. 1995. Approaches for assessing phylogenetic accuracy. Systematic Biology 44: 3–16.

——— and J. J. BULL. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Systematic Biology 42: 182–192.

——— and J. P. HUELSENBECK. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. Journal of Heredity 83: 189–195.

———, J. P. HUELSENBECK, and C. W. CUNNINGHAM. 1994a. Application and accuracy of molecular phylogenies. Science 264: 671–677.

———, ———, and D. L. SWOFFORD. 1994b. Hobgoblin of phylogenetics? Nature 369: 363–364.

HUELSENBECK, J. P. 1991a. Tree-length distribution skewness: An indicator of phylogenetic information. Systematic Zoology 40: 257–270.

———. 1991b. When are fossils better than extant taxa in phylogenetic analysis? Systematic Zoology 40: 458–469.

———. 1995. The performance of phylogenetic methods in simulation. Systematic Biology 44: 17–48.

——— and D. M. HILLIS. 1993. Success of phylogenetic methods in the four-taxon case. Systematic Biology 42: 247–264.

———, D. M. HILLIS, and R. JONES. In press. Parametric bootstrapping in molecular phylogenetics: Applications and performance. Pp. XX–XX in Molecular zoology: Strategies and protocols, eds. J. D. Ferraris and S. Palumbi. New York: Wiley-Liss.

KIM, J. and M. A. BURGMAN. 1988. Accuracy of phylogenetic-estimation methods under unequal evolutionary rates. Evolution 42: 596–602.

KORBER, B. T. M., R. M. FARBER, D. H. WOLPERT, and A. S. LAPEDES. 1993. Covariation of mutations in the V3 loop of human immunodefficiency virus type 1 envelope protein. Proceedings of the National Academy of Sciences, U.S.A. 90: 7176–7180.

LAMBOY, W. F. 1994. The accuracy of the maximum parsimony method for phylogeny reconstruction with morphological characters. Systematic Botany 19: 489–505.

LOSOS, J. B. 1994. An approach to the analysis of comparative data when a phylogeny is unavailable or incomplete. Systematic Biology 43: 117–123.

LYNCH, M. 1989. Phylogenetic hypotheses under the assumption of neutral quantitative-genetic variation. Evolution 43: 1–17.

MABEE, P. 1993. Phylogenetic interpretation of ontogenetic change: Sorting out the actual and artefactual in an empirical case study of centrarchid fishes. Zoological Journal of the Linnean Society 107: 175–291.

MADDISON, W. P. 1990. A method for testing the correlated evolution of two binary characers: Are gains and losses concentrated on certain branches of a phylogenetic tree? Evolution 44: 539–557.

NEI, M., F. TAJIMA, and Y. TATENO. 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. Journal of Molecular Evolution 19: 153–170.

PURVIS, A. and T. GARLAND, JR. 1993. Polytomies in comparative analyses of continuous characters. Systematic Biology 42: 569–575.

ROHLF, F. J. and M. C. WOOTEN. 1988. Evaluation of the restricted maximum-likelihood method for estimating phylogenetic trees using simulated allele-frequency data. Evolution 42: 581–595.

———, W. S. CHANG, R. R. SOKAL, and J. KIM. 1990. Accuracy of estimated phylogenies: Effects of tree topology and evolutionary model. Evolution 44: 1671–1684.

SANDERSON, M. J. and M. J. DONOGHUE. 1989. Patterns of variation in levels of homoplasy. Evolution 43: 1781–1795.

——— and J. J. DOYLE. 1992. Reconstruction of organismal and gene phylogenies from data on multigene families: Concerted evolution, homoplasy, and confidence. Systematic Biology 41: 4–17.

———, B. G. BALDWIN, G. BHARATHAN, C. S. CAMPBELL, C. VON DOHLEN, D. FERGUSON, J. M. PORTER, M. F. WOJCIECHOWSKI, and M. J. DONOGHUE. 1993. The growth of phylogenetic information and the need for a phylogenetic data base. Systematic Biology 42: 562–568.

STEVENS, P. F. 1991. Character states, morphological variation, and phylogenetic analysis: A review. Systematic Botany 16: 553–583.

SWOFFORD, D. L. and G. J. OLSEN. 1990. Phylogeny reconstruction. Pp. 411–501 in Molecular systematics, eds. D. M. Hillis and C. Moritz. Sunderland, Massachussetts: Sinauer Associates.

TATENO, Y., N. TAKEZAKI, and M. NEI. 1994. Relative efficiencies of the maximum-likelihood, neighbor-joining, and maximum-parsimony methods when substitution rate varies with site. Molecular Biology and Evolution 11: 261–277.

THIELE, K. 1993. The holy grail of the perfect character: The cladistic treatment of morphometric data. Cladistics 9: 275–304.

WHEELER, W. C. 1992. Extinction, sampling and molecular phylogenetics. Pp. 205–215 in Extinction and phylogeny, eds. M. Novacek and Q. Wheeler. New York: Columbia Univ. Press.

——— and R. L. HONEYCUTT. 1988. Paired sequence difference in ribosomal RNAs: Evolution and phylogenetic implications. Molecular Biology and Evolution 5: 90–96.

WIENS, J. J. 1993. Phylogenetic systematics of the tree lizards (genus Urosaurus). Herpetologica 49: 399–420.

———. 1995. Polymorphic characters in phylogenetic systematics. Systematic Biology 44: 482–500.

——— and P. T. CHIPPINDALE. 1994. Combining and weighting characters and the prior agreement approach revisited. Systematic Biology 43: 564–566.

——— and T. W. REEDER. 1995. Combining data sets with different numbers of taxa for phylogenetic analysis. Systematic Biology 44: 548–558.

WINTERBOTTOM, R. 1990. The Trimmatom nanus species complex (Actinopterygii, Gobiidae): Phylogeny and progenetic heterochrony. Systematic Zoology 39: 253–265.

ZARKIKH, A. and W.-H. LI. 1992a. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences: I. Four taxa with a molecular clock. Molecular Biology and Evolution 9: 1119–1147.

——— and ———. 1992b. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences: II. Four taxa without a molecular clock. Journal of Molecular Evolution 35: 356–366.

—JOHN J. WIENS[1] and DAVID M. HILLIS, Department of Zoology, The University of Texas, Austin, Texas 78712–1064

[1]Present Address: Carnegie Museum of Natural History, 4400 Forbes Avenue, Pittsburgh, Pennsylvania 15213–4080

Morphological Characters, Polytomies, and Homoplasy Indices: Response to Wiens and Hillis.

—ABSTRACT. This commentary responds to Wiens and Hillis' critique of a simulation study I conducted that examined the ability of maximum parsimony to find the true tree when morphological characters were used in the phylogenetic analysis. It rebuts all the main criticisms of