Total Evidence Or Taxonomic Congruence: Cladistics Or Consensus Classification

Arnold G. Kluge

Museum of Zoology, University of Michigan, Ann Arbor, Michigan, 48109, U.S.A.

Miyamoto and Fitch's (1995, *Syst. Biol.* 44: 64–76) verificationist arguments for taxonomic congruence are evaluated and found to be unconvincing. In particular, there is no logical connection between the truth of phylogenetic hypotheses and the independence of the sets of characters analysed for their consensus. Further, the character set partitions emphasized by Miyamoto and Fitch must be considered arbitrary, because they are based on untestable process assumptions. © 1998 The Willi Hennig Society

"There is no method of ascertaining the truth of a scientific hypothesis, i.e., no method of verification." "There is no method of ascertaining whether a hypothesis is 'probable', or probably true" Karl Popper (1992: 6).

INTRODUCTION

Miyamoto and Fitch (1995; hereafter referred to as MF; see also Lanyon, 1993) discuss a research program in phylogenetic inference, which I call *consensus classification*. The practice of consensus classification involves five essential elements: (1) *partitioning* data into subsets of characters (e.g. according to their origin from different genes); (2) justifying a particular partition with an *assumption(s)* concerning the

independence of the underlying process(es); (3) *weighting characters a priori*; and (4) forming a *consensus* of the *optimal and near optimal* phylogenetic hypotheses that result from the analysis of each partitioned subset of the data, and then forming a *consensus* of the consensuses obtained from the partitioned data sets. The core of consensus classification is taxonomic congruence, which includes partitioning data, and formulating consensus hypotheses of the results obtained from those separately analysed subsets of the data (Kluge, 1989). Element (5) or consensus classification, according to MF, is the measure of *truth* found in the consensus of the consensuses.

This paper begins with a brief account of phylogenetic systematics principles (Hennig, 1966; Farris, 1983), because it seems likely that the question of "cladistics or consensus classification" will be determined eventually by the scientific character of the two approaches. A restatement of total evidence and taxonomic congruence is included in this summary, because similarities and differences in these two practices are misrepresented by MF (for more detailed discussion see Kluge, 1989; Eernisse and Kluge, 1993; Jones et al., 1993; Kluge and Wolf, 1993). How consensus classification compares to cladistics is entertained in the next section of the paper, where each of the five elements is discussed. My chief concern is with MF's (p. 64) claim that "congruent [consensus] trees

obtained from analyses of independent data sets provide the best estimates of the true phylogeny for a group". I conclude by disputing the connection these authors make between "independent data sets" and "true phylogeny", that is, between *independence* and *truth*.

SOME PHYLOGENETIC SYSTEMATICS PRINCIPLES

The scientific character of cladistics can be understood largely in terms of the logical probability of Popperian testability, the probability of a hypothesis in light of its tests (Popper, 1968, 1972a,b, 1992; Farris, 1983; Kluge, 1997). Cladograms are tested in the severest manner possible, in relation to their empirical content, and the degree to which a cladistic hypothesis has withstood these attempts at refutation constitutes its degree of corroboration. Refutation and corroboration are the alternative results of testing. Explanatory power is directly related to degree of corroboration through their formal relationship to severity of test. Number of ad hoc hypotheses is also connected to explanatory power; however, that relationship is complementary. It is the least refuted cladogram which is the most parsimonious pattern of sister-group relationships, the hypotheses of relationships which has maximal content and which minimizes requirements for ad hoc propositions of homoplasy. As Farris (1983: 18) emphasized, the potential to maximize explanatory power, that is being able to provide an explanation for the similarity of congruent shared derived traits as due to inheritance, is merely a consequence of minimizing requirements for ad hoc hypotheses of homoplasy.

Advocacy for total evidence follows from the ideal that a scientist can do no better than explain the data. Also, in terms of testability, cladists use all of the relevant available synapomorphies, the total evidence, when testing a phylogenetic hypothesis, because a statement describing the results of *multiple* tests (especially if the tests are independent) "will be less probable than a statement describing only some of the tests" (Popper, 1992: 247–248)—a multiple test result being more improbable, and accordingly *more severe*, than its component tests.

Unlike total evidence, taxonomic congruence involves partitioning the relevant available evidence and the most parsimonious cladograms resulting from the separate analyses of those subsets of the data are then judged for similarity of relationships. And, according to MF (p. 67), "[a]s process partitions, the different evolutionary and biological properties of data sets make it more likely that the agreement among their topologies is the result of the true species phylogeny". As such, consensus classification is a neo-justificationist (frequency probabilist) research program, one in which the truth of a hypothesis is sought through induction (consensus).

Contrary to various statements by MF (e.g. p. 65 and elsewhere), consensus classification (taxonomic congruence) is similar to total evidence in that it uses cladistic parsimony to assess taxonomic relationships in each of the partitioned data sets. Thus, testability, the minimization of ad hoc requirements of homoplasy, and the assumption that the incongruent characters are independent, as parts of cladistic parsimony, can be viewed as relevant to *both* total evidence and taxonomic congruence. However, as pointed out by Kluge and Wolf (1993: 188, criticism 2), taxonomic congruence must then be considered an inconsistent methodology for hypothesizing relationships, because it involves the contradictory philosophies and operational criteria of cladistic parsimony and consensus, the former being refutationist in nature, the latter being one of verification (Kluge, 1997).

CLADISTICS OR CONSENSUS CLASSIFICATION

Cladistics and consensus classification have little in common when it comes to testing phylogenetic hypotheses, and nothing at all in common when it comes to choosing among possible alternatives and judging the reliability of one's choice. The terms *reliability*, *accuracy* and *precision* are often used in promoting taxonomic congruence, and may be misunderstood by some as having the same meaning. For the purposes of this discussion, I take accuracy to mean conformity to truth; precision pertains to exactness, being definite, clear, and unambiguous; and reliability concerns being worthy of confidence. In these terms,

precision has nothing to do necessarily with the truth (see also Hillis and Bull, 1993), and the importance of reliability in that connection depends on how confidence is understood (see below).

The Perceived Problem

Apparently, MF (p. 69) intend consensus classification for a narrowly defined class of data, nucleotide sequences, and where they believe incongruence between gene and species trees indicates significant amounts of homoplasy, sufficient to overwhelm the signal of homology that points to phylogenetic truth. There are two issues in relating gene trees to the species tree, one conceptual, the other operational. Conceptually, equating gene trees to the species tree is an error of over-reduction (e.g. Moore, 1995), just as is equating cell lineages to the species tree (e.g. Donoghue, 1985; K. de Queiroz and Donoghue, 1988). The particular significance of that error was illustrated by Frost and Kluge (1994: 266) with an analogy to historical individuality. As they pointed out, "[i]f evaluated at the scalar level of cell lineages, any coherent organismal corpus is 'paraphyletic', inasmuch as some cells of the body are more closely related to cells outside of the body … than to other cells within the body". Simply, gene and cell histories are non-transitive hierarchies, each having properties of its own which do not act equally at another hierarchical level, such as the species history.

To relate different kinds of scalar hierarchies, and scalar hierarchies to a specification hierarchy, such as a monophyletic taxonomy, requires an ontological "patch" (ad hocism, assumption; Frost and Kluge, 1994), and the discovery operation in question will determine the nature of that bridge. In phylogenetic inference, the "patch" might involve relating a gene tree (Moore, 1995), or set of congruent gene trees (MF), to the species tree, or the bridge between the two levels can involve assumptions having to do with nucleotide synapomorphies as potential homologues. In any case, there is a "patch", and the number and quality of the accompanying assumptions is important in choosing one discovery operation over another. In addition, whatever operation is undertaken, it must be able to evaluate the reliability of the gene and species trees.

Partitioning Data

The subdivision of data into subsets of characters (e.g. according to their origin from different genes), and their separate analysis with cladistic parsimony, defines consensus classification in operational terms. However, consider that there are a large number of possible ways in which a data set can be divided into two or more groups of characters, and it seems only reasonable to ask on what grounds one subdivision is to be preferred over any other, in the absence of knowledge of independent processes (see below; Kluge, 1989; Chippindale and Wiens, 1994). Is the separate analysis of different genes (e.g. γ^1-globin and 12S ribosomal RNA sequences) really to be determined by anything more than the differences in individual human interest, tradition or technology, which are the likely factors responsible for building those databases separately in the first place? Thus, the partitioning required by consensus classification is considered arbitrary.

There remains the concern for how partitioning relates to *data* (*characters*) *as evidence* for sister-group relationships (Kluge, 1989; see also Kluge and Wolf, 1993: 190, criticism 7). MF's (p. 66) assertion that "all of the available information is analyzed on a data set by data set basis" is beside the point, because that argument fails to make the distinction between synapomorphies generally (characters) and evidence in particular, just those incongruent synapomorphies (homoplasies) which count against a phylogenetic hypothesis (Kluge, 1997). Also too general is A. de Queiroz et al.'s (1995: 659–660; see also Huelsenbeck et al., 1996) claim that "[t]o qualify as a distinct class of evidence, characters in a data set must, in a statistical sense, be more similar to each other than they are to characters in other sets with respect to some property that affects phylogeny estimation by the given method". According to the principles of phylogenetic systematics, to qualify as different classes of evidence, characters (synapomorphies) are understood as homologues or non-homologues.

Thus, the critical issues for consensus classification are whether the homology and homoplasy classes of evidence can be related to partitioning, and whether consensus can effectively distinguish between hypotheses of homology and homoplasy. The distinction between gene and species trees (Goodman et al., 1979), and the possibility of those patterns being incongruent

(Pamilo and Nei, 1988; Doyle, 1992; Bull et al., 1993), have been considered relevant to these issues. For example, as A. de Queiroz et al. (1995: 664) conjectured, "separate analyses assume that the characters are independent estimators of the gene trees, not of the species tree, and consensus assumes that the different gene trees are not likely to differ from the species history in the same way". However, the absence of compelling empirical evidence for molecular characters being any more homoplasious than any other class of data (Sanderson and Donoghue, 1989; Donoghue and Sanderson, 1994; A. de Queiroz et al., 1995) suggests the need to partition nucleotide sequence data according to different genes is exaggerated. In any case, as the following discussion indicates, the consensus operation provides no real basis for distinguishing hypotheses of homology from homoplasy.

ASSUMPTIONS

Cladistics involves few assumptions, and its scientific character is due in no small part to its reliance on but a single auxiliary premise (background knowledge) having to do with process, namely descent with modification (Darwin, 1859: 420; Kluge, 1997). Further, cladists eschew special knowledge claims, like those concerning rates of evolution, organism fitness, and ancestral species. Such assumptions can carry a particularly heavy burden when they cannot be tested empirically (e.g. most recent common ancestral species), or when they can be judged only in light of the hypothesis which required the assumption in the first place (e.g. a constant rate of character evolution).

Bull et al. (1993) argued for partitioning characters according to process, yielding so-called process partitions, which they defined as subsets of the available data which are evolving under demonstrably different rules (MF, p. 68). Bull et al. took the position that significant differences in rules are demonstrated when data sets strongly support conflicting trees. In attempting to complete this line of argument for partitioning, MF (p. 68) assert that "[t]hese heterogeneities are not likely the result of stochastic errors alone but are more likely due to separate histories for the different sets of characters and/or to systematic errors and model failures in the phylogeny reconstructions for at least some of the data sets". MF go even further when arguing for the *independence* of process partitions. Thus, consensus classification includes an appeal to the special knowledge of partition independence, and such assumptions are absolutely critical to the proposition that taxonomic congruence assesses the accuracy of phylogenetic hypotheses.

MF (p. 68) list several examples of what they mean by independent processes, all of which require knowledge of genetic linkage, pleiotropy, and/or gene function. However, as MF admit, genetic processes are not the only guide to independence, and one is left to wonder what other classes of processes might be considered. What about those referred to as epigenetic? What about those processes that ultimately affect the phenotype, such as natural selection (or selective neutrality)? It is doubtful that any of these assumptions can really be tested empirically, and consensus classification must be considered very heavily burdened by having to assume detailed knowledge of a wide variety of possible processes. Perhaps, at best, the genetic processes responsible for nucleotide sequence partitions might be investigated empirically; however, in granting that possibility, consensus classification would be of limited application—it would not then be a general approach to phylogenetic inference.

Thus, unlike cladistics, consensus classification requires the extra assumption(s) concerning independence between data sets. However, matters are even worse. As pointed out by Barrett et al. (1991: 492), if dependent characters are to be combined with each other, it is not clear why the independence of characters is reason for keeping them apart. MF ignore this logical contradiction in taxonomic congruence (Kluge and Wolf, 1993: 192), and consensus classification must continue to be judged internally inconsistent in this regard.

Weighting Characters a Priori

It is becoming increasingly popular to treat molecular sequence data with weights derived a priori, before a hypothesis of sister-group relationships has been tested (Hillis et al., 1993: 475). These are not "observed" weights; they are expected weights, weights characterized by theory and methodology only indirectly related to a phylogenetic hypothesis (Mindell and

Thacker, 1996). Ordinarily, a verification justification accompanies a priori weighting, the signal (homology) to noise (homoplasy) ratio in the data supposedly being improved by the use of those weights. However, the refutationist epistemology of cladistics argues against the use of such a priori weighting, because the assumptions required to justify the expected weights add to background knowledge, which decreases the probability of a cladistic hypothesis in light of its tests. Adding to background knowledge is a "slippery slope", which ultimately ends in tautology (Kluge, 1998).

Consensus classification also involves another type of a priori character weighting, one that necessarily follows from the methodology of taxonomic congruence. As Kluge and Wolf (1993: 189, criticism 5) pointed out, a best-fitting hypothesis of sister-group relationships is sought for each partitioned set of characters, which is a form of equal weighting, and when the number of characters differs among the subsets of the data, as it usually does, the constituent characters are then unequally weighted. MF did not attempt to defend the fact "that the consensus procedure implies weightings of its own" (Barrett et al., 1991: 487), both equal and unequal.

Testability requires that each character in a data matrix provide an independent, potential disconfirming, test. Also, independent synapomorphies may be considered of equal weight in this sense. Differential weighting of characters according to their estimated or assumed independence has been explored by several authors (Shaffer, 1986; Wheeler and Honeycutt, 1988; Donoghue and Sanderson, 1992; Dixon and Hillis, 1993; Chippindale and Wiens, 1994; A. de Queiroz et al., 1995: 673), and there seems to be no technical reason why the amount of evidence for character independence that MF might use to justify process partitions cannot be incorporated into an a priori weighting formula, nor why such an approach would not be effective. The concept of weighting might include Doyle's (1992) treatment of a gene tree as a single multistate variable, and including it with other character data to obtain a globally most parsimonious hypothesis of species relationships. If weighing can be considered an operational substitute for the character partitioning operation, then there would appear to be no need for taxonomic congruence (MF, p. 66). Moreover, in the absence of taxonomic congruence, one is then free to pursue total evidence, where explanatory power is maximized, and the consensus of fundamental cladograms is avoided (Kluge and Wolf, 1993: figure 2).

Consensus

As noted above, in consensus classification, verification is sought through consensus. Even if consensus is viewed only as a basis for playing it safe, rendering a conservative hypothesis (Swofford, 1991), and not as a basis for assessing phylogenetic truth, it remains the antithesis of refutation, which seeks decisive outcomes (Popper, 1992). In addition, as MF proposed (see also Swofford, 1991), *optimal and near optimal* phylogenetic hypotheses that result from the analysis of each partitioned subset of the data may be used in forming a *consensus* of the consensuses obtained from the partitioned data sets. Unfortunately, MF fail to respond to the argument that there is no good way to judge how near to optimal a hypothesis must be to be included (Kluge and Wolf, 1993: 191). Thus, one is left to conclude that the decision is arbitrary.

Although character congruence is applied in the analysis of each partitioned data set in consensus classification, the effectiveness of such testing is necessarily less than it might be were the data to remain unpartitioned (Kluge, 1997). Also, there is no guarantee that explanatory power will be maximized on the consensus hypothesis of phylogenetic relationships (Miyamoto, 1985; Kluge, 1989). Moreover, there is the likely possibility that the consensus cladogram will be "positively at odds" with respect to the signal in the unpartitioned data set (Barrett et al., 1991: 489; for several examples see Chippindale and Wiens, 1994: table 1). However, matters are even worse, because MF (p. 70) advocate using the optimal *and* near optimal phylogenetic hypotheses obtained from the analysis of each partitioned subset (see also Swofford, 1991; Patterson et al., 1993: 178). As stated earlier, consensus classification must be judged internally inconsistent, because it involves the contradictory refutationist and verificationist philosophies and operational criteria.

The epistemological alternatives of refutation and verification necessarily have other important consequences for phylogenetic inference. For example, cladists and consensus classificationists exhibit

distinctly different attitudes toward homoplasy. A most parsimonious cladistic hypothesis describes the histories of all characters included in the analysis, those synapomorphies which may provide evidence for common ancestry (presumed homologues), as well as those features which cannot, because of their independent origins (presumed homoplasies). On the other hand, it is the nature of consensus classification, with its emphasis on combinable components, to focus on just those characters which can be interpreted as having had unique and unreversed histories—the discovery of independently evolved traits, and an assessment of their phylogenetic informativeness, being of little or no concern. Consensus classification and compatibility (clique) analysis are alike in this regard, and both must be considered incomplete verificationist approaches to phylogenetic inference (Farris and Kluge, 1979: 403). At the very least, consensus classification and compatibility analysis give the impression that homoplasy is not a historical pattern, and if not then they give the appearance of confounding ontology and epistemology (e.g. Turner and Zandee, 1995: 70).

Reliability

There are strongly held competing views on how reliability is assessed in phylogenetic inference. In general, cladists emphasize degree of corroboration (support) as the basis for hypothesis choice (e.g. Bremer, 1994). Also, as noted above, corroboration and refutation are simply alternative results of testing. The verificationist contrast to refutation is exemplified by MF (see also Lanyon, 1993), who emphasize induction, verification through consensus, and unabashedly state their claims in terms of truth (not just verisimilitude!). For example, MF (p.64), and elsewhere (Miyamoto and Cracraft, 1991; Miyamoto et al., 1994: 236; see also Hillis and Bull, 1993: 485), contend that taxonomic congruence "is part of the age-old tradition in science of *hypothesis testing with new independent information. It remains the most familiar way for practicing systematists to test the accuracy [truth] of their phylogenetic conclusions*" (my italics). The traditional form of hypothesis testing referred to by MF is induction, "the collection and (statistical) tabulation of instances, especially confirming instances" (Popper, 1992: 256).

However, MF fail to mention that it is highly debatable whether induction can assess accuracy or provide a decisive basis for choosing among hypotheses (see epigraph), and that being a "tradition" carries no weight in arguments pro or con induction.

According to MF, reliability is some function of consensus—a consensus is formed of the optimal and near optimal phylogenetic hypotheses that result from the analysis of each partitioned subset of the data, and then a consensus is formed of the consensuses obtained from the partitioned data sets. As underscored earlier, consensus classification assumes that the different data sets analysed for consensus are independent process partitions. And, again according to MF (p.64), it is the consensus derived from such differently defined data sets that "are expected to converge onto the *true species phylogeny*" (my italics; see also Miyamoto and Cracraft, 1991; Lanyon, 1993; Sheldon and Bledsoe, 1993; Miyamoto et al., 1994; A. de Queiroz et al., 1995).

MF do not discuss how confidence in a consensus hypothesis is to be measured exactly, nor how partitioning data provides a "more powerful test" than a total evidence analysis. However, confidence in taxonomic congruence is suggested in MF's assertion that "[a]s process partitions, the different evolutionary and biological properties of data sets make it more likely that the agreement among their topologies is the result of the true species phylogeny" (p.67; see also Prager and Wilson, 1978; McKitrick, 1985; Hillis, 1987, 1995; Miyamoto and Cracraft, 1991; Sheldon and Bledsoe, 1993: 256–257). How convincing can this claim be, however, when total evidence analyses give, as they usually do, different and more fully explained results from those produced by taxonomic congruence? Moreover, those who endorse taxonomic congruence do not all agree as to the nature of consensus. For example, Hillis (1995: 11; see also Swofford, 1991) stated that "it is important *not* to interpret consensus trees as estimates of phylogenies but rather simply as statements about areas of agreement among trees" (my italics). According to Hillis and Bull (1993; see also Felsenstein, 1985), it is repeatability, not accuracy, which is at issue. In any case, shifting the focus to "areas of agreement" still leaves open the questions of how to optimize that criterion, (e.g. which consensus method to use), and what might justify that criterion (see Kluge and Wolf, 1993: 188, criticism 3). The recent literature on component analysis and three-taxon statements suggests that

finding satisfactory answers to these two questions will be difficult (for recent review see Kluge, 1994).

Two simple examples illustrate how difficult it is to connect independence to accuracy in phylogenetic inference. For the sake of argument in the first example, assume rooted cladogram (A,B)C to be true, that is to say that taxa A and B share a more recent common ancestor than either does with C. Further assume that there are two or more characters (indefinitely many if you wish), or sets of characters, whose synapomorphy distribution 110 fits exactly that true phylogeny. In terms of deductive logic, there are two premises in this example, (A,B)C and 110, and although those premises are perfectly consistent some extra assumption is required to link them together *causally*—a causation that would lead to the conclusion that congruent synapomorphies are indeed homologues (Farris, 1983). Contrary to what MF assert, the bridge is not made by measuring the independence of characters, either as pairs, or as larger sets. Of course, failure of the assumption of character independence can significantly alter one's choice of the most parsimonious cladogram (Farris, 1983); however, that effect by itself does not measure the truth of those hypothesized sister-group relationships.

Character compatibility analysis provides another example, and one where the absence of a reasonable argument for bridging has contributed to the method's failure to be widely accepted in phylogenetic inference. Although logic may dictate that "[a]*ll characters that support true evolutionary history must be members of the same clique*" (Meacham, 1980: 156), that is a long way from being able to ascribe truth to any particular set of compatible characters, even the set (clique) with the largest number of characters. In fact, all that one can conclude logically from compatibility analysis is that one or the other, or both, of two *incompatible* characters must be homoplasious. Adding the condition of independence to each of the pairwise compatible characters in the largest clique does not make a bridge to homologues and the truth—there remains a leap of faith.

Some argument, assumption, or theory is necessary to be able to bridge the assumption of independence to the truth, and MF fail to make that connection. The only thing "more" that taxonomic congruence offers over total evidence is the requirement for special knowledge. Such special knowledge claims will simply not carry the day given the coherence and

generality of cladistics, of which total evidence is a part. Maximizing explanatory power may not be a guarantee of accuracy, "but then accuracy as it pertains to knowing the *truth* is not an obsession of cladists" (Kluge, 1995: 77).

ACKNOWLEDGEMENTS

REFERENCES

Barrett, M., Donoghue, M. J., and Sober, E. (1991). Against consensus. *Syst. Zool.* **40**, 486–493.

Bremer, K. (1994). Branch support and tree stability. *Cladistics* **10**, 295–304.

Bull, J. J., Huelsenbeck, J. P., Cunningham, C. W., Swofford, D. L., and Waddell, P. J. (1993). Partitioning and combining data in phylogenetic analysis. *Syst. Biol.* **42**, 384–397.

Chippindale, P. T., and Wiens, J. J. (1994). Weighting, partitioning, and combining characters in phylogenetic analysis. *Syst. Biol.* **43**, 278–287.

Darwin, C. (1859). "On the Origin of Species". 1964 facsimile reprint of first edition. Harvard University Press, Cambridge, Massachusetts.

De Queiroz, A., Donoghue, M. J., and Kim, J. (1995). Separate versus combined analysis of phylogenetic evidence. *Ann. Rev. Ecol. Syst.* **26**, 657–681.

De Queiroz, K., and Donoghue, M. J. (1988). Phylogenetic systematics and the species problem. *Cladistics* **4**, 317–338.

Dixon, M. T., and Hillis, D. M. (1993). Ribosomal RNA secondary structure: Compensatory mutations and implications for phylogenetic analysis. *Mol. Biol. Evol.* **10**, 256–267.

Donoghue, M. J. (1985). A critique of the biological species concept and recommendations for a phylogenetic alternative. *Bryologist* **88**, 172–181.

Donoghue, M. J., and Sanderson, M. J. (1992). The suitability of molecular and morphological evidence in reconstructing plant phylogeny. *In* "Molecular Systematics of Plants" (P. S. Soltis, D. E. Soltis, and J. J. Doyle, Eds.), pp. 340–368. Chapman and Hall, New York.

Donoghue, M. J., and Sanderson, M. J. (1994). Complexity and homology in plants. *In* "Homology: The Hierarchical Basis of Comparative Biology" (B. Hall, Ed.), pp. 393–421. Academic Press, San Diego, California.

Doyle, J. J. (1992). Gene trees and species trees: Molecular systematics as one-character taxonomy. *Syst. Bot.* **17**, 144–163.

Eernisse, D., and Kluge, A. G. (1993). Taxonomic congruence versus total evidence, and amniote phylogeny inferred from fossils, molecules, and morphology. *Mol. Biol. Evol.* **10**, 1170–1195.

Farris, J. S. (1983). The logical basis of phylogenetic analysis. *In* "Advances in Cladistics" (N. I. Platnick and V. A. Funk, Eds.), pp. 7–36. Vol. 2, Columbia University Press, New York.

Farris, J. S. (1995). Conjectures and refutations. *Cladistics* **11**, 105–118.

Farris, J. S., and Kluge, A. G. (1979). A botanical clique. *Syst. Zool.* **28**, 400–411.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**, 783–791.

Frost, D. R., and Kluge, A. G. (1994). A consideration of epistemology in systematic biology, with special reference to species. *Cladistics* **10**, 259–294.

Goodman, M., Czelusniak, J., Moore, G. W., Romero-Herrera, A. E., and Matsuda, G. (1979). Fitting the gene lineage into its species lineage: A parsimony strategy illustrated by cladograms constructed from globin sequences. *Syst. Zool.* **28**, 132–168.

Hennig, W. (1966). "Phylogenetic Systematics". University of Illinois Press, Urbana, Illinois.

Hillis, D. M. (1987). Molecular versus morphological approaches to systematics. *Ann. Rev. Ecol. Syst.* **18**, 23–42.

Hillis, D. M. (1995). Approaches for assessing phylogenetic accuracy. *Syst. Biol.* **44**, 3–16.

Hillis, D. M., and Bull, J. J. (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* **42**, 182–192.

Hillis, D. M., Allard, M. W., and Miyamoto, M. M. (1993). Analysis of DNA sequence data: Phylogenetic inference. *Methods Enzymol.* **224**, 456–487.

Huelsenbeck, J. P., Bull, J. J., and Cunningham, C. W. (1996). Combining data in phylogenetic analysis. *Trends Ecol. Evol.* **11**, 152–158.

Jones, T. R., Kluge, A. G., and Wolf, A. J. (1993). When theories and methodologies clash: A phylogenetic reanalysis of the North American ambystomatid salamanders (Caudata: Ambystomatidae). *Syst. Biol.* **42**, 92–102.

Kluge, A. G. (1989). A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Syst. Zool.* **38**, 7–25.

Kluge, A. G. (1994). Moving targets and shell games. *Cladistics* **10**, 403–413.

Kluge, A. G. (1995). Parsimony. *Herpetol. Rev.* **26**, 76–77.

Kluge, A. G. (1997). Testability and the refutation and corroboration of cladistic hypotheses. *Cladistics* **13**, 81–96.

Kluge, A. G. (1998). Sophisticated falsification and research cycles: Consequences for differential character weighting in phylogenetic systematics. *Zool. Scripta* (in press).

Kluge, A. G., and Wolf, A. J. (1993). Cladistics: What's in a word. *Cladistics* **9**, 183–199.

Lanyon, S. M. (1993). Phylogenetic frameworks: Towards a firmer foundation for the comparative approach. *Biol. J. Linn. Soc.* **49**, 45–61.

McKitrick, M. C. (1985). Monophyly of the Tyrannidae (Aves): Comparison of morphology and DNA. *Syst. Zool.* **34**, 35–45.

Meacham, C. A. (1980). Phylogeny of the Berberidaceae with an evaluation of classifications. *Syst. Bot.* **5**, 149–172.

Mindell, D., and Thacker, C. (1996). Rates of molecular evolution: Phylogenetic issues and applications. *Ann. Rev. Ecol. Syst.* **27**, 279–303.

Miyamoto, M. M. (1985). Consensus cladograms and general classifications. *Cladistics* **1**, 186–189.

Miyamoto, M. M., and Cracraft, J. (1991). Phylogenetic inference, DNA sequence analysis, and the future of molecular systematics. *In* "Phylogenetic Analysis of DNA Sequences" (M. M. Miyamoto, and J. Cracraft, Eds.), pp. 3–17. Oxford University Press, New York.

Miyamoto, M. M., and Fitch, W. M. (1995). Testing species phylogenies and phylogenetic methods with congruence. *Syst. Biol.* **44**, 64–76.

Miyamoto, M. M., Allard, M. W., Adkins, R. M., Janecek, L. L., and Honeycutt, R. L. (1994). A congruence test of reliability using linked mitochondrial DNA sequences. *Syst. Biol.* **43**, 236–249.

Moore, W. S. (1995). Inferring phylogenies from mtDNA variation: Mitochondrial-gene trees versus nuclear-gene trees. *Evolution* **49**, 718–726.

Pamillo, P., and Nei, M. (1988). Relationships between gene trees and species trees. *Mol. Biol. Evol.* **5**, 568–583.

Patterson, C., Williams, D. M., and Humphries, C. J. (1993). Congruence between molecular and morphological phylogenies. *Ann. Rev. Ecol. Syst.* **24**, 153–188.

Popper, K. R. (1968). "The Logic of Scientific Discovery". Harper and Row, New York.

Popper, K. R. (1972a). "Conjectures and Refutations: The Growth of Scientific Knowledge". Routledge and Kegan Paul, London.

Popper, K. R. (1972b). "Objective Knowledge: An Evolutionary Approach". Oxford University Press, Oxford.

Popper, K. R. (1992). "Realism and the Aim of Science". Routledge, London.

Prager, E. M., and Wilson, A. C. (1978). Construction of phylogenetic trees for proteins and nucleic acids: Empirical evaluation of alternative matrix methods. *J. Mol. Evol.* **11**, 129–142.

Sanderson, M. J., and Donoghue, M. J. (1989). Patterns of variation in levels of homoplasy. *Evolution* **43**, 1781–1795.

Shaffer, H. B. (1986). Utility of quantitative genetic parameters in character weighting. *Syst. Zool.* **35**, 124–134.

Sheldon, F. H., and Bledsoe, A. H. (1993). Avian molecular systematics, 1970s to 1990s. *Ann. Rev. Ecol. Syst.* **24**, 243–278.

Swofford, D. L. (1991). When are phylogeny estimates from molecular and morphological data incongruent? *In* "Phylogenetic Analysis of DNA Sequences" (M. M. Miyamoto, and J. Cracraft, Eds.), pp. 295–333. Oxford University Press, New York.

Turner, H., and Zandee, R. (1995). The behaviour of Goloboff's tree fitness measure *f*. *Cladistics* **11**, 57–72.

Wheeler, W. C., and Honeycutt, R. L. (1988). Paired sequence differences in ribosomal RNAs: Evolution and phylogenetic implications. *Mol. Biol. Evol.* **5**, 90–96.