

CLADISTICS: WHAT'S IN A WORD?

Arnold G. Kluge¹ and Alan J. Wolf¹

¹ Museum of Zoology and Department of Biology, University of Michigan,
Ann Arbor, Michigan 48109, U.S.A.

Received for publication 17 July 1992; accepted 4 March 1993

Abstract—Cladistics has changed considerably with the availability of new methods and sources of data, and the increasing realization that cladograms are relevant to all manner of historical questions. Criticisms of, and justifications for, consensus hypotheses in phylogenetic inference are reviewed. The conclusion is overwhelmingly against taxonomic congruence which deliberately seeks consensus propositions. The total evidence approach is not so burdened. A preference for suboptimal cladograms is also criticized, as is the protocol for mapping characters of special interest onto a phylogenetic hypothesis derived from other evidence. The bootstrap and jackknife resampling techniques are questioned because their underlying assumptions are violated and they are sensitive to character frequencies. These findings suggest that cladistics is being redefined in ways that contradict the practices and principles responsible for its pre-eminence in phylogenetic inference.

Introduction

Cladistics has become the predominant paradigm in phylogenetic inference. Further, there can be little doubt that the past success of cladistics is attributable to the generality and coherence of its practices and principles relating to monophyly, parsimony and special similarity. However, other factors are contributing currently to its increasing popularity. Simply, the methodological and philosophical limits of the discipline are now being relaxed. In effect, cladistics is becoming all things to all people, which we believe diminishes significantly the critical content of the paradigm. Thus, we examine issues relevant to the changing definition of cladistics, in particular taxonomic congruence, the use of suboptimal hypotheses, the nature of character independence, and congruence and the statistical confidence ascribed to the bootstrap and jackknife resampling techniques.¹ We hope that our review will make systematists more aware of the changing boundaries of phylogenetic systematics, and that additional critical comment will follow. While our position on these topics can be interpreted as “doctrinaire”, we believe our study has the potential to strengthen historical inference by exposing arguments antithetical to the principles of cladistics and those *ad hoc* methods that only bear the label, not the meaning, of phylogenetic systematics.

Taxonomic Congruence

Seeking congruent classifications of the same group of organisms is an old strategy in systematics, and one that has had various objectives. For example, such an approach was discussed in detail by Sokal and Sneath (1963: 85–91) as a way of testing the non-specificity hypothesis (Farris, 1971). Shortly thereafter (Sneath and Sokal, 1973: 66), taxonomic congruence became a tool in the arguments

¹ The three-taxon transformation of Nelson and Platnick (1991), also a part of this re-examination, is discussed separately (Kluge, 1993).

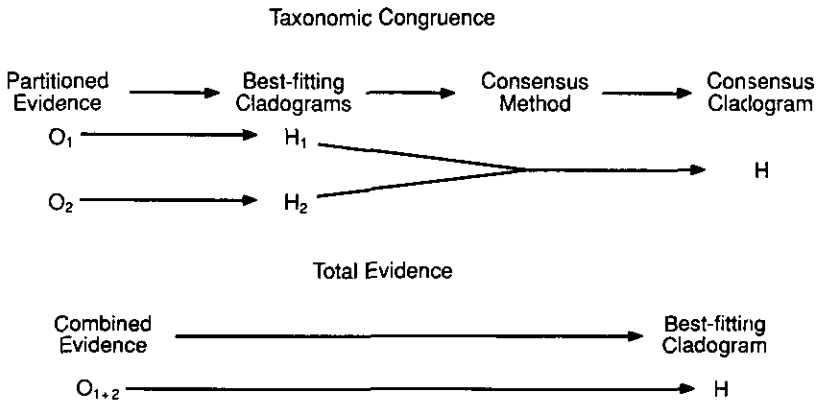


Fig. 1. Taxonomic congruence and total evidence are compared in terms of the operations involved in analyzing partitioned and combined relevant available evidence, respectively. The former approach necessarily employs a consensus method. Relevant evidence is that which is potentially phylogenetically informative (i.e. synapomorphies). Ideally, total evidence can be taken to mean all of the evidence available. Practically speaking, it is all of the unpartitioned evidence at hand. After Eernisse and Kluge (1993).

between pheneticists and phylogeneticists over whose methods led to more stable classifications (see review by Kluge, 1989a). It was in the latter context that Mickevich (1978: 143) developed her definition: "Taxonomic congruence is the degree to which classifications of the same organisms postulate the same groupings. When the classifications compared are based on different sources of information, congruence [consensus] provides a measure of the degree to which the classifications of the organisms remain stable as various lines of evidence are considered".

In the following discussion of taxonomic congruence, we focus on different data sets (*sensu* Kluge, 1989a), rather than alternative methods of phylogenetic inference (Crisci, 1984; Bledsoe and Raikow, 1990). It is in that specific context that we consider both the operation and the various justifications for taxonomic congruence, including that of stability which pheneticists also continue to use as a basis for recommending their classifications (Sokal, 1986: 424).

Ordinarily, the methodology of taxonomic congruence involves three steps (Kluge, 1989a): (1) evidence is partitioned into "different *types* of data" (e.g. Shaffer et al., 1991: 284; our italics); (2) best-fitting hypotheses of relationships are obtained from each set of data (so-called fundamental cladograms are produced); and (3) a consensus of those topologies is formulated [Fig. 1]. In contrast, total evidence uses character congruence to find the best-fitting phylogenetic hypothesis for an unpartitioned set of synapomorphies, which ideally is all of the relevant available data [Fig. 1]. Equally parsimonious propositions, referred to as secondary cladograms by Kluge (1989a: fig. 3), can also be discovered with total evidence, and consensus methods have been used to summarize those alternatives (Fig. 2).

Data quality is not sufficient to distinguish taxonomic congruence from total evidence, because the same characters can be involved. Thus, the pejorative claim that the total evidence approach involves relatively poorly studied characters is not on point.

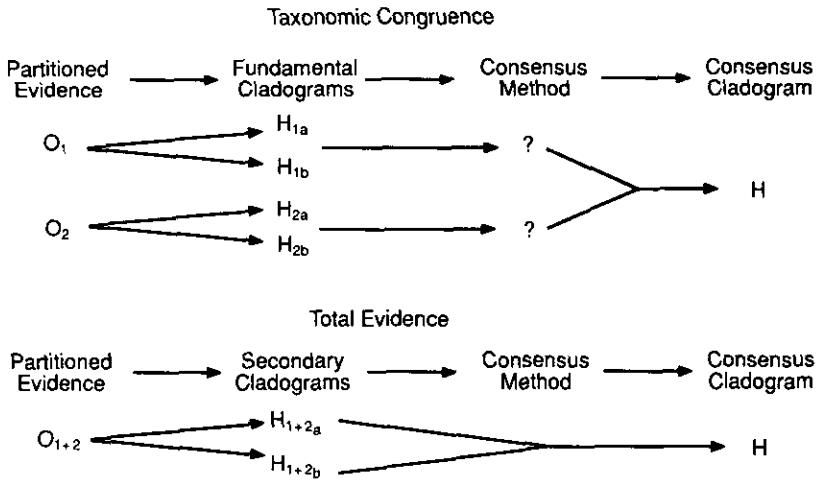


Fig. 2. Equally parsimonious, fundamental and secondary, cladograms can result from taxonomic congruence and total evidence approaches, respectively. The alternative hypotheses provided by secondary cladograms can be analyzed with a consensus method (as illustrated) or additional cycles of research (see text for further explanation; Kluge, 1991). It is unclear how to analyze suites of equally parsimonious fundamental cladograms. After Eernisse and Kluge (1993).

The criticisms we set forth below are organized into those against constructing consensus hypotheses in general (1–3), whatever their basis (fundamental or secondary cladograms), and those pertaining just to consensuses of fundamental cladograms (4–7). While some of the problems we discuss are of a general nature, there is nonetheless an important distinction between taxonomic congruence and total evidence. The method of taxonomic congruence involves an *advocacy* for different types of data, and a consensus of fundamental cladograms is touted as something more—for example, being a *more probable* hypothesis of species relationships (Nelson, 1979: 1, 5–7, 8, 13; see, however, Mickevich, 1978: 155; Rohlf, 1982: 139–140; and Farris, 1991: 91; see also discussion of independence below). On the contrary, total evidence attempts to find the maximally described and explained hypothesis for all of the available evidence, and that objective can be realized when more than one secondary cladogram is identified by: (1) seeking additional data; (2) re-examining incongruent characters and states for erroneous estimation of polarity, definition and description (*sensu* the research cycles of Kluge, 1991); (3) using a *posteriori* (successive) weighting to test the “cladistic reliability” of the incongruent characters (Farris, 1969: 374; Carpenter, 1988; Platnick et al., 1991); and/or (4) employing a consilience of inductions, where choice among equally parsimonious hypotheses is based on how well they are explained by extrinsic evidence, such as provided by earth history and the phylogenetic hypothesis of one or more groups of coevolving organisms (e.g. hosts/parasites; Kluge, 1983: 172–173). We do not recommend solving the problem of equally most parsimonious hypotheses by sanitizing the data, by substituting a more inclusive group name for the lineages in the original matrix whose relationships cannot be decisively delimited, because this simply hides the problem. In any case, criticisms 1–3 do not *necessarily* burden total evidence as they do taxonomic congruence.

One may be *forced* to apply consensus to secondary cladograms when new sources of evidence, intrinsic and extrinsic, are unavailable, research cycles have been exhausted, successive weighting is ineffective, and saving journal space is demanded. One may even use consensus hypotheses to underscore the differences in explanatory power between total evidence and taxonomic congruence. As our criticisms will show, such uses of consensus are benign compared to their scientifically pernicious effect in taxonomic congruence.

CRITICISM I

As Farris (1979, 1980, 1982, 1983, 1985b, 1989: 107; Farris and Kluge, 1985, 1986) argued convincingly, biological classifications represent theories, and their usefulness is to be found in how well they describe and explain the natural processes that give rise to observations. Thus, phylogenetic hypotheses and their corresponding taxonomies have descriptive and explanatory power, and Miyamoto (1985: 187; see also Carpenter, 1988; Kluge, 1989a; Maddison, 1989; and Barrett et al., 1991) was one of the first to call attention to the fact that consensus cladograms of different fundamental topologies "cannot faithfully produce parsimonious solutions from multiple character sets, because they cannot resolve conflict and ambiguity according to evidence". Less-resolved topologies require more synapomorphies be explained as homoplasies (however, see below). Of course, the more the fundamental, or secondary, cladograms differ, the more the consensus hypothesis will be unresolved, and the greater will be the loss of descriptive and explanatory power.

Recently, Swofford (1991: 311; see also Anderberg and Tehler, 1990) discussed this particular criticism, and challenged the relevance of descriptive and explanatory power in biological classification. He asserted that loss of such power is not really a problem for taxonomic congruence because "[p]olytomous nodes on a consensus tree do not indicate simultaneous cladogenetic events", they are only "areas of uncertain resolution". To Swofford, "[c]onsensus trees are simply statements about areas of agreement among trees; they should not be interpreted as phylogenies". However, other historical biologists (e.g. Maddison, 1989) have considered polytomies, which can result from either absence of evidence or incongruent synapomorphies (Kluge, 1989b), to be either multiple-speciation events *or* evidence of uncertain relationships. Moreover, descriptive and explanatory power is important in investigating the nature of speciation (Wiley, 1981: 43–56), just as it is in trying to understand the basis for the incongruent characters responsible for uncertain relationships (Rieppel, 1989).

Swofford's (1991: 329) conviction that "an ambiguous solution [polytomy] that contains the truth is, in many situations, preferable to an unambiguous solution that is wrong" would appear to justify his view of consensus hypotheses. But, such a stance cannot be a compelling argument for how to practice cladistics, or any other science for that matter, because a completely unresolved hypothesis would then certainly contain the truth. Moreover, one might argue that completely resolved phylogenetic hypotheses are to be preferred, because "[t]he more a theory forbids, the better it is" (Popper, 1965: 36). In summary, polytomies might be claimed to be more accurate, but they are surely less precise, and in the absence of knowing the truth we are left with precision.

Similarly, consensus may be viewed as a way of reaching the *safest* conclusion possible—"by restricting one's final hypothesis to the points about which several competing hypotheses agree, one appears to run a reduced risk of being mistaken" (Barrett et al., 1991: 486). However, safety, like stability (predictivity) in classification (Kluge, 1989a: 7–8), is not the ideal pursued by cladists, nor the criterion by which they judge their labors. Likewise, strict consensus may be the most *conservative* method, because it "instructs one to find those hypotheses about which *all* the best trees agree" (Barrett et al., 1991: 486; our italics), but that does not provide a compelling justification for its use in phylogenetic inference. In fact, one could argue just as well that bold (Popper, 1968: 281), not safe or conservative, hypotheses mark cladistics as a science, because those kinds of propositions emphasize disconfirmation, rather than verifiability (Platnick and Gaffney, 1977, 1978a,b).

Hillis (1987) offered yet another view on the usefulness of polytomies from taxonomic congruence. While he (p. 35) endorsed the position that "the best estimate of phylogeny and the best estimate of character evolution are represented in the analysis of the combined data sets", it was claimed that "[c]lassifications [taxonomies] are best based on information in common among multiple data sets (i.e., consensus trees)". Hillis (see also Anderberg and Tehler, 1990) effectively attributed different descriptive and explanatory powers to cladograms and hierarchies of names, and we doubt that solution will be acceptable to many phylogeneticists. Surely, cladists should object.

Lastly, Cracraft and Helm-Bychowski (1991: 187; see also Cracraft and Mindell, 1989) attempted to make a case for taxonomic congruence with notions of reliability and informativeness. They stated: "Ideally, 'reliability' is best thought of as a parsimony problem and consequently should be judged in terms of congruence of data. 'Informativeness', likewise, is also a parsimony problem. The two concepts are related. A given tree, calculated to be the most-parsimonious, or best-fit, for some set of data, can be taken to be a reliable estimate of the true phylogenetic history of the taxa if most-parsimonious trees for other, independent data are congruent with it. A given sample of [DNA] sequence can be judged phylogenetically informative if it corroborates the most-parsimonious tree derived from other data. This reciprocal duality between the structure of data and what is inferred from them manifests the application of parsimony (and congruence analysis) to hypotheses at two different levels of inference: that of the data (character homology or hypotheses of synapomorphy), and that of the phylogenetic hypotheses themselves". Cracraft and Helm-Bychowski (1991: 187) then went on to assess informativeness and reliability in terms of *stability*, as judged by the "cladistic signal across partitioned subsets of the data. That is, to the extent to which clades are corroborated by these different subsets, the data will be judged informative and the phylogenetic hypotheses themselves reliable". However, as noted above, *stability per se* is not the goal of cladistics, and therefore we must conclude that reliability and informativeness (*sensu* Cracraft and Helm-Bychowski) provide no justification for taxonomic congruence. To be sure, "phylogenetic *hypotheses* are better supported when additional independent data are provided that lead to the same conclusion as the initial data set", but "[p]hylogenetic hypotheses should not be stable under the addition of new characters, if those characters in fact suffice to disconfirm genealogical conjectures that were accepted earlier" (Sober, 1988: 141–142; see also Kluge, 1989a: 7–82).

We believe Cracraft and Helm-Bychowski have conflated topological identity with evidential corroboration.

CRITICISM 2

Both taxonomic congruence and total evidence approaches may employ one rule, parsimony, in the discovery of fundamental and secondary cladograms, respectively, and another rule, consensus, in summarizing equally parsimonious hypotheses. Parsimony and consensus cannot be the same in so far as they involve different justifications and methodologies, and they can lead to contradictory results. Cracraft and Helm-Bychowski (1991:187) provided no evidence for their assertion that "informativeness [consensus] . . . is also a parsimony problem". Of course, character congruence is operationally the same as consensus when characters are analyzed two at a time, as in compatibility analysis. However, that is not the ordinary sense in which character congruence is understood, especially when one considers that congruence is maximized on the best-fitting hypothesis for *all* of the evidence. Thus, to include parsimony *and* consensus in the same decision-making sequence, relevant to discovering species relationships, is inconsistent, viz. a particular criterion (e.g. parsimony) is and is not the basis for discovering sister-group relationships for a given set of data. Certainly, logical inconsistency contributed to compatibility analysis' failure to become widely accepted as a method for cladogram construction—the desire for complete resolution of relationships required that previously incompatible characters be employed at a later stage in the analysis (Farris and Kluge, 1979: 402). We predict the same future for those methods that necessarily seek consensus hypotheses.

CRITICISM 3

There are many methods for forming a consensus topology, and such methods do not always give the same result when applied to a particular combination of data sets. The problem is that choice among methods appears to be independent of any reasoned biological considerations (Kluge, 1989a: 9), and disagreements are likely to be common. There are even different opinions concerning the strict consensus method's popularity in phylogenetic inference. Some claim that its appeal is due to "simplicity of interpretation" (Swofford, 1991: 297), while others judge it to be "the most conservative" (Barrett et al., 1991: 487). It seems likely that choice of methods and adjudication of different results will, at best, involve only technical considerations of the algorithms (e.g. see review by Swofford, 1991). Thus, most choices will be arbitrary, and such a basis for action has nothing to recommend it in science. Arbitrary choice among countless methods was certainly one of the factors that doomed phenetics.

CRITICISM 4

In addition to the loss of descriptive and explanatory power, Barrett et al. (1991; see also Brooks, 1988) demonstrated that taxonomic congruence can actually provide a positively misleading result (Nelson's, 1993, interpretation aside; see Barrett et al., 1993). As can be seen from Barrett's et al. example (redrafted here as

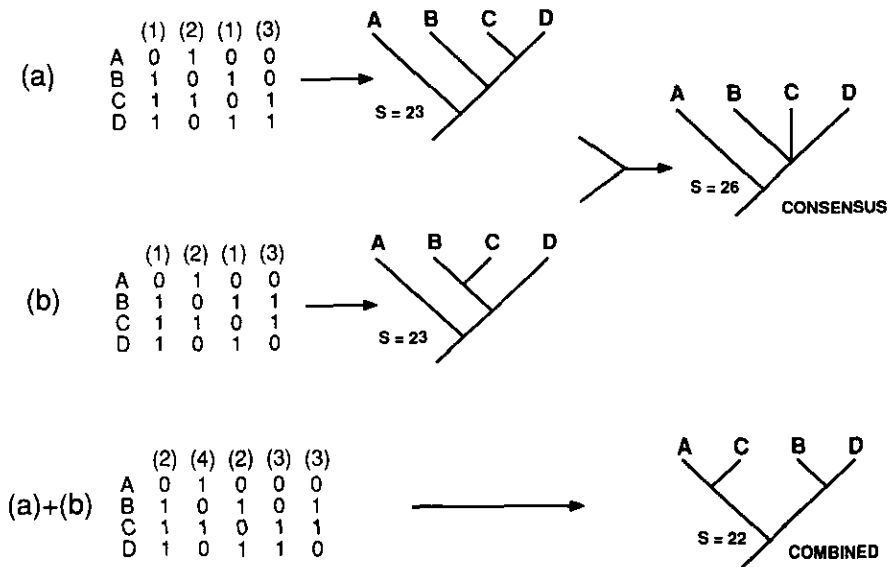


Fig. 3. The results of taxonomic congruence can be positively misleading with respect to those obtained from an analysis of the total evidence (example from Barrett et al., 1991: fig. 1). Two sets of seven binary characters each, (a) and (b), describe four taxa, A–D. State 0 is plesiomorphic. The character distributions are identical in the two data sets, except for the synapomorphies which delimit the (C, D) and (B, C) clades, (a) and (b), respectively. Unique best-fitting cladograms obtained from each data set form the basis for the strict consensus hypothesis. A different topology is obtained from the total evidence, (a) + (b). Note the sister-group relationships of taxon C in the four cladograms. S is the number of evolutionary events required to explain the evidence; the smaller the number the better fitting is the corresponding hypothesis. See text for further explanation.

Fig. 3), the fundamental cladograms based on data sets (a) and (b), and their strict consensus hypothesis, are inconsistent with the best-fitting branching pattern for the pooled evidence, (a) + (b). The sister-group (C, D) is specified by (a), (B, C) by (b), (B, C, D) by their consensus and (A, C) by (a) + (b). Moreover, the total evidence analysis provides a better-fitting hypothesis (fewer number of steps, $S = 22$), than does either of the fundamental cladograms ($S = 23$), or their strict consensus ($S = 26$), for the combined data. Barrett et al. (p. 488) argued that other consensus methods are not immune to this problem because “[o]f all the different methods, strict consensus is most likely to yield a result that is consistent with the tree produced from a combined data set”. They concluded (p. 488) “that all consensus methods can endorse trees that contradict the tree obtained from the pooled data”, and that (p. 491) “the consensus tree cannot be regarded as the best (or ‘safest’) inference to make from the available data”.

CRITICISM 5

Taxonomic congruence involves equal weighting of data sets, and therefore the constituent characters are unequally weighted when the sizes of the data sets differ (Miyamoto, 1985; see also Cracraft and Mindell, 1989; and Kluge, 1989a). As Barrett et al. (1991: 487) noted, the real problem is not differential character weighting, but “that the consensus procedure implies weightings of its own, and these may be difficult to defend”.

CRITICISM 6

Another issue peculiar to taxonomic congruence concerns the approach one should take in summarizing a suite of equally parsimonious cladograms obtained from one data set with a suite of equally parsimonious cladograms retrieved from another body of evidence. Do you take a consensus of each suite of fundamental cladograms, and a consensus of those consensuses (Colless, 1981), or a consensus of the two sets of branching patterns (Swofford, 1991: 319)? Or, as per Mickevich's (1980; Schuh and Farris, 1981: 340) method, "the congruence of each tree in one suite with each tree in the other suite is measured, and the congruence of the suites is taken as the maximum value of these pairwise comparisons"? While Schuh and Farris (1981) have argued convincingly why Colless' (1981) method obscures disagreement between data sets, choice among the other two approaches remains an open question (see also Swofford, 1991: 317). Moreover, that the arguments pro and con these alternatives can be affected by which consensus method is chosen (see criticism 3 above) would appear to mean that an ambiguous decision will always obtain.

CRITICISM 7

Partitioning evidence according to different types of data is fundamental to taxonomic congruence, but in so doing one presupposes that classes of evidence (e.g. molecular and morphological) actually exist in nature (Kluge, 1991). We have yet to hear a convincing argument for such a reality, as mind-independent categories, by those who search for historical patterns of relationships (e.g. Swofford, 1991: 328–329). To appreciate better the arbitrary nature of partitioned evidence consider that allozymes can be defined as molecular, and proteins as morphological; however, such a distinction belies their closeness ontogenetically, and separating them into these classes obscures their lack of independence. Just as cladists reject non-monophyletic higher taxa because of their artificiality, so must they question artificial subdivisions of evidence because there is no reason to believe those definitions have discoverable boundaries.

While systematists interested in historical patterns and processes may study classes of characters, such as molecular and morphological, for reasons of tradition and technology, such subdivisions do not provide a reasonable basis for claiming the existence of classes of evidence. Evidence is evidence, and to cladists evidence is a character match judged with character congruence to be a homologue (Patterson, 1982). There is nothing in the prescription for studying phenotypic and genotypic variation that requires it to be organized into multiple character sets in order to formulate phylogenetic hypotheses.

Suboptimal Hypotheses

Cracraft and Helm-Bychowski's (1991: 184) "methodological strategies" and Swofford's (1991: 318) "multifaceted approach" for inferring phylogeny have one obvious thing in common, a preference for suboptimal hypotheses (see also Hillis and Dixon, 1989). Cracraft and Helm-Bychowski's promotion of such a questionable scientific philosophy lies in their attempt to assess the "phylogenetic signal"

(pp. 189, 215), as a measure of stability; however, as we have already argued, stability *per se* is not the goal of phylogenetic systematics. Swofford's (1991: 329) justification for suboptimal hypotheses fares no better in our opinion. He asserted "that an ambiguous solution [a consensus of non-minimal trees] that contains the truth is, in many situations, preferable to an unambiguous solution that is wrong". But, as we noted above, this can hardly be considered a scientifically-sound recommendation. Further, neither of these groups of authors blunted the criticism that the choice of a set of suboptimal cladograms is arbitrary, and of course, as Hillis and Dixon (1989) and Swofford (1991) observed, consensus of different sets of near minimum length trees are often different. Still further, the continuous nature of tree length distributions (e.g. see Hillis, 1991: fig. 13.1; and Swofford, 1991: fig. 14.7) indicates that a set of suboptimal hypotheses could reasonably include every possible cladogram for which there is some evidence, and once the decision is made to pursue non-minimal length cladograms it does not end effectively until the least optimal is included.² While such a generous course of action may avoid the charge of arbitrariness, there is nothing else to recommend it. Lastly, we emphasize that choosing sets of nearly equally most parsimonious hypotheses means that some consensus method has to be employed in order to discover the commonalities. Again, there is more arbitrariness, as can be seen in the variety of consensus methods employed by Cracraft and Helm-Bychowski (1991) and Swofford (1991).

Character Independence and Character Congruence

Character independence is an assumption that all phylogeneticists have to make, whether they practice taxonomic or character congruence. That two or more characters are independent can be taken to mean that they evolved at different places and/or times within a species lineage (along a branch or interval of a cladogram; Kluge, 1990: fig. 1).³ The reason character independence must be considered in phylogenetic inference concerns the fact that amount of evidence is used in choosing among alternative hypotheses of history, and therefore "each synapomorphy in the character matrix is assumed to count as a *separate* piece of evidence, viz. has the potential to confirm, or disconfirm, taxonomic relationships independent of all other synapomorphies considered" (Kluge, 1989a: 11).

There are two different contexts in which the assumption of character independence can be investigated. The first, and easiest, is to eliminate all redundancies and correlations that occur in the descriptions of the different characters (e.g. Smith, 1992). Those characters whose descriptions have passed such tests may be termed "logically" independent (Farris, 1983: 20; see however, Swofford, 1991: 327).

² This notion gave rise to the "slippery slope" metaphor employed by A. G. Kluge in his talk, "The principle of total evidence: A maxim in phylogenetic inference", presented in the Society of Systematic Zoology symposium, December, 1989, "Recent Advances in Phylogenetic Studies of DNA Sequences", organized by J. Cracraft and M. M. Miyamoto.

³ E. Sober (pers. comm.) has suggested a more general definition of independence, one in keeping with his maximum likelihood justification of parsimony (Sober, 1988). Namely, two characters are independent if the probability of one of them changing on a given branch is independent of the other's changing. Simply, probabilistic independence means that the probability that both characters change is the product of the probabilities that each changes.

The second context can be much more difficult because an intraspecific historical hypothesis may be necessary to judge whether or not the evolution of two characters occurred at the same place and time (Kluge, 1991: figs 1–2). Of course, congruent characters that are separated by a cladistic event can be assumed to be independent, but such *prima facie* evidence is unavailable for judging synapomorphies at the same level of generality. Unfortunately, even our best-supported phylogenetic propositions do not provide details of tokogenetic relationships (Hennig, 1966), such that those synapomorphies can be judged as having had an independent origin. Furthermore, the actual processes responsible for non-independent evolution cannot be read from even the most detailed patterns of organism relationships.

Nonetheless, Swofford (1991: 328) claimed that the assumption of independence “is usually much more plausible for characters in different sets than for characters within the same data set”. While we acknowledge that he may be correct in unusual cases, such as comparisons of phylogenetic hypotheses generated with clonally inherited mtDNA or clonally inherited Y-chromosome sequences,⁴ there is no necessary connection between “types” of characters and independence (see also Nelson, 1979). Exactly why is it any more “plausible” to consider molecular and morphological classes of characters to be independent than, say, skeletal versus visceral characters? More importantly, the assumption of character independence does not take on any greater significance between data sets than it does within a data set. As Barrett et al. (1991: 492) stated, “[i]f independent morphological (or molecular) characters are to be combined with each other, it is hard to see why the ‘independence’ of morphological and molecular characters is a reason for keeping them apart”. We do not accept Swofford’s (1991: 328) claim that “[t]he corroboration provided by the independent estimates would not be available if the data sets were combined”.⁵

The importance of history in the fields of behavior and ecology is now firmly established, as it is generally where past events relevant to processes are judged in the context of a cladistic hypothesis. However, with this new-found relevance has come a misunderstanding largely, if not entirely, promoted by systematists. It has become common practice in phylogenetic inference to separate the characters used to build a hypothesis of sister-group relationships (character matrix X) from those characters whose histories are to be judged in terms of that hypothesis [character set Y (Y can be one or more characters)]. In particular, we refer to Brooks and McLennan’s (1991: 63; see also Coddington, 1988: 7, 12, 20) “cardinal rule: *never use the characters that are part of the evolutionary hypothesis under investigation to build your phylogenetic tree*. Rather, these characters should be mapped onto an existing tree”. While their rule took various forms (e.g. p. 134: “never bias your analysis by using the ecological information you want to study to build your phylogenetic tree”), its basic message is that characters are not to

⁴ Even in relatively clear-cut examples such as this, exceptions may exist in which there may be biparental inheritance of mtDNA (Hoeh et al., 1991). Thus, non-independence between paternally- and maternally-based phylogenies is *theoretically* possible.

⁵ We believe the notion of corroboration should be reserved for the hierarchically correlated evidence hypothesized by character congruence (Farris, 1969). The correlations provided by taxonomic congruence will not be the same, to the extent the fundamental cladograms are different, and therefore corroboration and consensus are not strictly equivalent.

be involved as evidence of history which are themselves to be interpreted historically.

More specific arguments than "bias" have been used to justify the mapping strategy. For example, as already noted above, Swofford (1991: 328) claimed that the assumption of independence "is usually much more plausible for characters in different datasets than for characters within the same dataset". Also, Coddington (1988: 7) asserted that, to eliminate the taint of "circularity", the morphological and functional characters involved in the historical study of adaptation should not be used in formulating the cladograms on which the adaptive hypothesis is reflected. According to Brooks and McLennan (1991: 134), to follow the rule is a way to avoid changing a hypothesis of character history when different outgroups are used (pp. 63 and 141). In addition, Pagel and Harvey (1988; see also Harvey and Pagel, 1991) argued that the effects of adaptation and phylogenetic constraint must be partitioned, and practicing the rule may be seen as a way of accomplishing that separation.

We are not persuaded by any of these arguments. That of character independence is not compelling, because the assumption of independence is made whether or not a character is part of the total evidence or mapped on secondarily (see above). Further, homologues give a cladogram meaning as a phylogenetic hypothesis, and operationally character congruence (finding the best-fitting cladogram for a matrix of synapomorphies) ultimately shapes our propositions of homology and non-homology. Alter the pattern of taxonomic relationships and the process explanation of some synapomorphy, inheritance due to common ancestry or independent evolution, will likely change. Thus, the evolutionary history of synapomorphies will be judged finally in terms of the best-fitting cladogram, we believe, for the entire, unpartitioned, body of relevant evidence (e.g. Platnick et al., 1991). The act of "mapping" a subset of characters on a cladogram derived from other evidence, even in a parsimoniously optimal manner, to avoid the charge of circularity, or for whatever reason, is not congruence in the aforementioned sense because mapped characters have no effect on the phylogenetic hypothesis because they are not independent of it. Both the homology of such characters and the parts of history those synapomorphies mark are effectively ignored by mapping (Mickey, 1982: 476-477).

Of course, the potential for producing the incorrect cladistic hypothesis is magnified when a large data set is mapped on a cladogram, as is the case with Swofford's (1991: 317) newly proposed measure of incongruence (I_M). Mickey and Farris' (1981) alternative method for partitioning the total incongruence into within- and between-data set components does not have that problem (Kluge, 1989a). The fact that Mickey and Farris method includes a total evidence measure of incongruence is actually an asset, not a "difficulty" (*contra* Swofford, 1991: 316).

As an aside, we see no reason why historical behaviorists and ecologists should not fully participate in formulating the historical propositions on which they judge character transformations of special interest. To use merely the efforts of cladists for the purposes of mapping only removes the behaviorist and ecologist from taking full responsibility for their conclusions and maintains the artificial separation of research disciplines. The larger research program of comparative biology (Rieppel, 1988), in which systematics is central, will not be allowed to grow naturally under those circumstances.

Resampling Techniques

Many systematists attach considerable significance to the bootstrap and jackknife resampling techniques,⁶ but we doubt that an interpretation of statistical significance is warranted.⁷ Our position rests on the fact that the methods in question make several assumptions, and that most, if not all, of those premises appear to be violated. For example, Felsenstein (1985) acknowledged that the following assumptions apply to bootstrapping: (1) the elements in the distribution are "independently and identically distributed (i.i.d.)" (p. 785); (2) the sample observations are collected without bias by the investigator (viz. randomly); (3) the sample of observations must be very large, such that each possible value will be represented in the same proportion that occurs in the parameter distribution—in other words, resampling from the original sample of observations with replacement will be the same as sampling from the parameter distribution; (4) the algorithm used to form a phylogenetic hypothesis from a data matrix actually finds the branching pattern of character covariation that best fits the data, and that in rejecting the null-hypothesis of random permutation, the most parsimonious estimate is the phylogeny (Sanderson, 1989).

To justify the "i.i.d." assumption, that each character evolved independently and is identically distributed, Felsenstein (1985: 784) assumed that evolution is stochastic. Under such a model of evolution, a character would appear by chance among and within the branches of the underlying phylogeny. Thus, according to Felsenstein, "[e]ach character is then a random sample from a distribution of all possible configurations of characters". However, Felsenstein acknowledged that characters do not evolve independently of each other, and he admitted (p. 785: see also Bermingham and Avise, 1986: 943) that this violation of the "i.i.d." assumption is "the most serious challenge to the use of bootstrap methods". That concern must also hold for the jackknife procedure (Felsenstein, 1988: 551), as well as permutation tests, such as the permutation tail probability (PTP) of Faith and Cranston (1991: 21–23), because they make the same assumption. Regarding the absence of investigator bias in sampling characters from the underlying distribution, Felsenstein (1985: 785) further admitted that systematists may not sample characters at random. We think it is more accurate to say the assumption is never met (see also J. S. Farris, *In*: Werdelin 1989: 192–193). As Sanderson (1989: 115) noted, "[t]his assumption is clearly at odds with many systematists' professed intention of selecting only the more evolutionarily conservative characters for study". Usually, the sample size of characters, the third assumption, is far too small to justify using bootstrapping. While it is not clear how large the sample of characters must be, it seems safe to say that it is smaller than it could be when only parts of the relevant available evidence are analyzed, as with taxonomic congruence. Given that most systematists investigate phylogenetic relationships with a limited data set (relative to that which is available in the comparative biology literature), the assumption of large sample size must be understood to be violated generally (e.g. Shaffer et al., 1991). Also, it should be emphasized, to ascribe 95%

⁶ Some journals, such as *Molecular Biology and Evolution*, in their instructions to authors, even "encourage" bootstrapping cladograms.

⁷ This is aside from the fact that the number of replications needed for precise estimation is almost always too small (Hedges, 1992), including, for example, the 1000 employed by Shaffer et al. (1991).

confidence to a particular clade that group must be delimited by *at least* three congruent synapomorphies (Felsenstein, 1985; Sanderson and Donoghue, 1989). And, the fact that the resampling method only provides a confidence value for each clade, not the entire cladogram (Felsenstein, 1988: 551; Sanderson, 1989), means that no part of a perfectly-fitting, ladder-like cladistic hypothesis, where each group is described by two congruent synapomorphies, can be significant at the 95% level. Lastly, as Sanderson (1989: 115) pointed out, "there seems little point in estimating sampling distributions from an underlying distribution of characters if the estimator itself [the algorithm for finding best-fitting hypotheses of relationships], when applied to that distribution, provides misleading results". Thus, we must assume minimally that the algorithm employed finds the best-fitting hypothesis(es). When heuristic algorithms are used to analyze highly incongruent data this assumption may be violated.

In addition, Felsenstein's (1985) bootstrap method is sensitive to the frequency of the characters in the data matrix (J. S. Farris, pers. comm.; Faith and Cranston, 1991: 21; Carpenter, 1992). The inclusion of autapomorphies actually lowers confidence limits, and many such uninformative characters are observed in allozyme and nucleotide sequence data. Also, the fact that the number of synapomorphies at one level of generality can diminish the significance of less generally shared derived states clearly calls into question Felsenstein's (1985: 784) stochastic model of evolution and his formulation of the bootstrap method. Even adding invariant characters can change the bootstrap "statistic".⁸

From this brief review of the assumptions made by the bootstrap and jackknife resampling procedures, and the prospect that most, if not all, will be violated, we find conclusions derived from bootstrapped (and jackknifed) confidence limits to be without basis. Further, these methods cannot, on pain of circular reasoning, be used to investigate character independence because characters must be "regarded as independently evolved on the same phylogeny" (Felsenstein, 1988: 551; *contra* Shaffer et al., 1991). Still further, the bootstrap and jackknife, like the matrix permutation approaches of Archie (1989) and Faith and Cranston (1991), suffer from not having a clear relationship to corroboration, which is central to cladistics (Carpenter, 1992).

Discussion

Our survey of taxonomic congruence, suboptimal hypotheses, character independence and congruence, and the bootstrap and jackknife resampling techniques, identifies practices and principles that differ significantly from those ordinarily understood to be Hennigian. Obviously, if cladists accept these additions then the tenets of cladistics will be changed and its limits redefined accordingly. However, before that happens, we urge that the more general consequences of such actions be considered carefully. For example, we believe the methodological strategies of taxonomic congruence and the use of suboptimal hypotheses involve many decisions arrived at without reason, and it will not be possible to argue that cladistics is coherent and logically consistent if those practices become a routine

⁸ Goloboff (1991; see also Kallersjö et al., 1992) should be consulted for a more lengthy discussion of the affect of character frequencies.

part of phylogenetic systematics. In addition, we observe that these strategies, as well as the bootstrap and jackknife resampling techniques, have more advocates among those who consider their particular type of data as special, such as molecular biologists. Thus, we would have to conclude that cladistics will become special-purpose, rather than remain general, with the approval of these strategies and techniques. Special purpose translates to *ad hoc*.

Also, we urge careful consideration of the argument for taxonomic congruence, and *consensus in particular*—that it allows all possible evidence to be compared, such as provided by character and molecular distance data. But, we must ask, what can be learned about phylogeny from a comparison of taxonomic relationships based on, say, morphology and immunological or DNA annealing (hybridization) data sets? When the fundamental topologies obtained from such sets confirm, there remains the empirically difficult issue of independence, as discussed above. Even less satisfactory, when those topologies disconfirm, there appears to be no empirical basis for evaluating the differences—cladistic rules of evidence do not apply to distance data (J. S. Farris, *In: Werdelin*, 1989: 192–193; Kluge, 1989a: 21). Given that “molecular distances seldom provide much useful information on phylogenetic relationships” (Farris, 1981, 1985a: 67), one is left to wonder how the analysis of such data can be used as a justification for taxonomic congruence and consensus (Cracraft and Mindell, 1989: 399).

Finally, we imagine that some of the popularity of the aforementioned methodological strategies and resampling techniques, and the assumption of independence in the context of taxonomic congruence and the cardinal rule of Brooks and McLennan (1991), derives from the belief that phylogenetic inference is hypothetico-deductive (e.g. Nelson and Platnick, 1984: 143–144), or at least that it should be. Even the uses to which some might put cladograms, such as “testing” adaptation (Coddington, 1988),⁹ are presented as hypothetico-deductive. But, this ignores an alternative, that cladistics, and its uses, may be an abductive enterprise (Sober, 1988). We suggest that the limits of phylogenetic systematics will be clarified considerably when cladists understand how their knowledge claims are made (Rieppel, 1988; Panchen, 1992).

Acknowledgments

The authors wish to thank the following for their critical input, both as regards ideas and more effective ways to communicate our concerns: J. M. Carpenter, D. J. Eernisse, J. S. Farris, W. R. Fink, D. R. Frost, T. R. Jones, M. Källersjö, J. S. Keogh, M. McKittrick, R. W. Murphy and E. Sober. These acknowledgments do not necessarily mean that our published paper is a consensus of their opinions. This research was supported by National Science Foundation grant BSR-8822656 to AGK.

REFERENCES

- ANDERBERG, A. AND A. TEHLER. 1990. Consensus trees, a necessity in taxonomic practice. *Cladistics* 6: 399–402.

⁹ Bearing in mind that, according to Popper (1965: 36), “[e]very genuine *test* of a theory is an attempt to falsify it, or to refute it. Testability is falsifiability”.

- ARCHIE, J. W. 1989. A randomization test for phylogenetic information in systematic data. *Syst. Zool.* 38: 239–252.
- BARRETT, M., M. J. DONOGHUE AND E. SOBER. 1991. Against consensus. *Syst. Zool.* 40: 486–493.
- BARRETT, M., M. J. DONOGHUE AND E. SOBER. 1993. Crusade? A reply to Nelson. *Syst. Biol.* 42: 216–217.
- BERMINGHAM, E. AND J. AVISE. 1986. Molecular zoogeography of freshwater fishes in the southeastern United States. *Genetics* 113: 939–965.
- BLEDSE, A. H. AND R. J. RAIKOW. 1990. A quantitative assessment of congruence between molecular and nonmolecular estimates of phylogeny. *J. Mol. Evol.* 30: 247–259.
- BROOKS, D. R. 1988. Macroevolutionary comparisons of host and parasite phylogenies. *Ann. Rev. Ecol. Syst.* 19: 235–259.
- BROOKS, D. R. AND D. A. MCLENNAN. 1991. *Phylogeny, Ecology and Behaviour: A Research Program in Comparative Biology*. Univ. Chicago Press, Chicago.
- CARPENTER, J. M. 1988. Choosing among multiple equally parsimonious cladograms. *Cladistics* 4: 291–296.
- CARPENTER, J. M. 1992. Random cladistics. *Cladistics* 8: 147–153.
- CODDINGTON, J. A. 1988. Cladistic tests of adaptational hypotheses. *Cladistics* 4: 3–22.
- COLLESS, D. H. 1981. Predictivity and stability in classifications: Some comments on recent studies. *Syst. Zool.* 30: 325–331.
- CRACRAFT, J. AND K. HELM-BYCHOWSKI. 1991. Parsimony and phylogenetic inference using DNA sequences: Some methodological strategies. *In: M. M. Miyamoto and J. Cracraft (eds). Phylogenetic Analysis of DNA Sequences*. Oxford Univ. Press, New York, pp. 184–220.
- CRACRAFT, J. AND D. P. MINDELL. 1989. The early history of modern birds: A comparison of molecular and morphological evidence. *In: B. Fernholm, K. Bremer and H. Jörnvall (eds). The Hierarchy of Life*. Elsevier, Amsterdam, pp. 389–403.
- CRISCI, J. V. 1984. Taxonomic congruence. *Taxon* 33: 233–239.
- EERNISSE, D. J. AND A. G. KLUGE. 1993. Taxonomic congruence versus total evidence, and the phylogeny of amniotes inferred from fossils, molecules and morphology. *Mol. Biol. Evol.* 10(6).
- FAITH, D. P. AND P. S. CRANSTON. 1991. Could a cladogram this short have arisen by chance alone?: On permutation tests for cladistic structure. *Cladistics* 7: 1–28.
- FARRIS, J. S. 1969. A successive approximation approach to character weighting. *Syst. Zool.* 18: 374–385.
- FARRIS, J. S. 1971. The hypothesis of nonspecificity and taxonomic congruence. *Ann. Rev. Ecol. Syst.* 2: 277–302.
- FARRIS, J. S. 1979. The information content of the phylogenetic system. *Syst. Zool.* 28: 483–519.
- FARRIS, J. S. 1980. The efficient diagnoses of the phylogenetic system. *Syst. Zool.* 29: 386–401.
- FARRIS, J. S. 1981. Distance data in phylogenetic analysis. *In: V. A. Funk and D. R. Brooks (eds). Advances in Cladistics, Vol. 1*. New York Bot. Garden, New York, pp. 3–23.
- FARRIS, J. S. 1982. Simplicity and informativeness in systematics and phylogeny. *Syst. Zool.* 31: 413–444.
- FARRIS, J. S. 1983. The logical basis of phylogenetic analysis. *In: N. I. Platnick and V. A. Funk (eds). Advances in Cladistics, Vol. 2*. Columbia Univ. Press, New York, pp. 7–36.
- FARRIS, J. S. 1985a. Distance data revisited. *Cladistics* 1: 67–85.
- FARRIS, J. S. 1985b. The pattern of cladistics. *Cladistics* 1: 190–201.
- FARRIS, J. S. 1988. Hennig86, Version 1.5. Port Jefferson Station, New York.
- FARRIS, J. S. 1989. Entropy and fruit flies. *Cladistics* 5: 103–108.
- FARRIS, J. S. 1991. Excess homoplasy ratios. *Cladistics* 7: 81–92.
- FARRIS, J. S. AND A. G. KLUGE. 1979. A botanical clique. *Syst. Zool.* 28: 400–411.
- FARRIS, J. S. AND A. G. KLUGE. 1985. Parsimony, synapomorphy, and explanatory power: A reply to Duncan. *Taxon* 34: 130–135.
- FARRIS, J. S. AND A. G. KLUGE. 1986. Synapomorphy, parsimony, and evidence. *Taxon* 35: 298–306.
- FELSENSTEIN, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39: 783–791.

- FELSENSTEIN, J. 1988. Phylogenies from molecular sequences: Inference and reliability. *Ann. Rev. Genetics* 22: 521–565.
- GOLOBOFF, P. A. 1991. Random data, homoplasy and information. *Cladistics* 7: 395–406.
- HARVEY, P. H. AND M. D. PAGEL. 1991. *The Comparative Method in Evolutionary Biology*. Oxford Univ. Press, New York.
- HEDGES, S. B. 1992. The number of replications needed for accurate estimation of the bootstrap *P* value in phylogenetic studies. *Mol. Biol. Evol.* 9: 366–369.
- HENNIG, W. 1966. *Phylogenetic Systematics*. Univ. Illinois Press, Chicago.
- HILLIS, D. M. 1987. Molecular versus morphological approaches to systematics. *Ann. Rev. Ecol. Syst.* 18: 23–42.
- HILLIS, D. M. 1991. Discriminating between phylogenetic signal and random noise in DNA sequences. In: M. M. Miyamoto and J. Cracraft (eds). *Phylogenetic Analysis of DNA Sequences*. Oxford Univ. Press, New York, pp. 278–294.
- HILLIS, D. M. AND M. T. DIXON. 1989. Vertebrate phylogeny: Evidence from 28S ribosomal DNA sequences. In: B. Fernholm, K. Bremer and H. Jörnvall (eds). *The Hierarchy of Life*. Elsevier, Amsterdam, pp. 355–367.
- HOEH, W. R., K. H. BLAKELY AND W. M. BROWN. 1991. Heteroplasmy suggests limited biparental inheritance of *Mytilus* mitochondrial DNA. *Science* 251: 1488–1490.
- KALLERSJÓ, M., J. S. FARRIS, A. G. KLUGE AND C. BULT. 1992. Skewness and permutation. *Cladistics* 8: 275–287.
- KLUGE, A. G. 1983. Cladistics and the classification of the great apes. In: R. L. Ciochon and R. S. Corruccini (eds). *New Interpretations of Ape and Human Ancestry*. Plenum Press, New York, pp. 151–177.
- KLUGE, A. G. 1989a. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Syst. Zool.* 38: 7–25.
- KLUGE, A. G. 1989b. Metacladistics. *Cladistics* 5: 291–294.
- KLUGE, A. G. 1990. Species as historical individuals. *Biol. Phil.* 5: 417–431.
- KLUGE, A. G. 1991. Boine snake phylogeny and research cycles. *Misc. Publ. Mus. Zool., Univ. Mich.* (178): 1–58.
- KLUGE, A. G. 1993. Three-taxon transformation in phylogenetic inference: ambiguity and distortion as regards explanatory power. *Cladistics* 9: 246–259.
- MADDISON, W. P. 1989. Reconstructing character evolution on polytomous cladograms. *Cladistics* 5: 365–377.
- MICKEVICH, M. F. 1978. Taxonomic congruence. *Syst. Zool.* 27: 143–158.
- MICKEVICH, M. F. 1980. Taxonomic congruence: Rohlf and Sokal's misunderstanding. *Syst. Zool.* 29: 162–176.
- MICKEVICH, M. F. 1982. Transformation series analysis. *Syst. Zool.* 31: 461–478.
- MICKEVICH, M. F. AND J. S. FARRIS. 1981. The implications of congruence in *Menidia*. *Syst. Zool.* 30: 351–370.
- MIYAMOTO, M. M. 1985. Consensus cladograms and general classifications. *Cladistics* 1: 186–189.
- NELSON, G. 1979. Cladistics analysis and synthesis: Principles and definitions, with a historical note on Adanson's *Familles des Plantes* (1763–1764). *Syst. Zool.* 28: 1–21.
- NELSON, G. 1993. Why crusade against consensus? A reply to Barrett, Donoghue, and Sober. *Syst. Biol.* 42: 215–216.
- NELSON, G. AND N. I. PLATNICK. 1984. Systematics and evolution. In: M.-W. Ho and P. T. Saunders (eds). *Beyond Neo-Darwinism*. Academic Press, New York, pp. 143–158.
- NELSON, G. AND N. I. PLATNICK. 1991. Three-taxon statements: A more precise use of parsimony? *Cladistics* 7: 351–366.
- PAGEL, M. D. AND P. H. HARVEY. 1988. Recent developments in the analysis of comparative data. *Quart. Rev. Biol.* 63: 413–440.
- PANCHEN, A. L. 1992. *Classification, Evolution, and the Nature of Biology*. Cambridge Univ. Press, New York.
- PATTERSON, C. 1982. Morphological characters and homology. In: K. A. Joysey and A. E. Friday (eds). *Problems of Phylogenetic Reconstruction*. Academic Press, New York, pp. 21–74.

- PLATNICK, N. I., J. A. CODDINGTON, R. R. FORSTER AND C. E. GRISWOLD. 1991. Spinneret morphology and the phylogeny of haplogyne spiders (Araneae, Araneomorphae). *Amer. Mus. Novitates* (3016): 1-73.
- PLATNICK, N. I. AND E. S. GAFFNEY. 1977. Systematics: A Popperian perspective. *Syst. Zool.* 26: 360-365.
- PLATNICK, N. I. AND E. S. GAFFNEY. 1978a. Evolutionary biology: A Popperian perspective. *Syst. Zool.* 27: 137-141.
- PLATNICK, N. I. AND E. S. GAFFNEY. 1978b. Systematics and the Popperian paradigm. *Syst. Zool.* 27: 381-388.
- POPPER, K. R. 1965. *Conjectures and Refutations: The Growth of Scientific Knowledge*. Harper and Row, New York.
- POPPER, K. R. 1968. *The Logic of Scientific Discovery*. Harper and Row, New York.
- RIEPEL, O. 1988. *Fundamentals of Comparative Biology*. Birkhauser Verlag, Boston.
- RIEPEL, O. 1989. Character incongruence: Noise or data? *Abh. Naturwiss. Ver. Hamburg* (28): 53-62.
- ROHLE, F. J. 1982. Consensus indices for comparing classifications. *Math. Biosoc.* 59: 131-144.
- SANDERSON, M. J. 1989. Confidence limits on phylogenies: The bootstrap revisited. *Cladistics* 5: 113-129.
- SANDERSON, M. J. AND M. J. DONOGHUE. 1989. Patterns of variation in levels of homoplasy. *Evolution* 43: 1781-1795.
- SCHUH, R. T. AND J. S. FARRIS 1981. Methods for investigating taxonomic congruence and their application to the Leptopodomorpha. *Syst. Zool.* 30: 331-351.
- SHAFFER, H. B., J. M. CLARK AND F. KRAUS. 1991. When molecules and morphology clash: A phylogenetic analysis of North American ambystomatid salamanders (Caudata: Ambystomatidae). *Syst. Zool.* 40: 284-303.
- SMITH, G. R. 1992. Introgression in fishes: Significance for palaeontology, cladistics, and evolutionary rates. *Syst. Biol.* 41: 41-57.
- SNEATH, P. H. A. AND R. R. SOKAL. 1973. *Numerical Taxonomy*. W. H. Freeman and Co., San Francisco.
- SOBER, E. 1988. *Reconstructing the Past: Parsimony, Evolution, and Inference*. MIT Press, Cambridge, Massachusetts.
- SOKAL, R. R. 1986. Phenetic taxonomy: Theory and methods. *Ann. Rev. Ecol. Syst.* 17: 423-442.
- SOKAL, R. R. AND P. H. A. SNEATH. 1963. *Principles of Numerical Taxonomy*. W. H. Freeman and Co., San Francisco.
- SWOFFORD, D. L. 1991. When are phylogeny estimates from molecular and morphological data incongruent? *In*: M. M. Miyamoto and J. Cracraft (eds). *Phylogenetic Analysis of DNA Sequences*. Oxford Univ. Press, New York, pp. 295-333.
- WERDELIN, L. 1989. We are not out of the woods yet—A report from a Nobel Symposium. *Cladistics* 5: 192-200.
- WILEY, E. O. 1981. *Phylogenetics: The Theory and Practice of Phylogenetic Systematics*. J. Wiley & Sons, New York.

