REPLY Taxic Homology = Overall Similarity

A. G. Kluge* and J. S. Farris†

*Division of Reptiles and Amphibians, Museum of Zoology, The University of Michigan, Ann Arbor, Michigan 48109-1079 USA; and †Molekylarsystematiska laboratoriet, Naturhistoriska riksmuseet, Box 50007 SE-104 05 Stockholm, Sweden

Accepted May 20, 1999

Modified three-taxon analysis (m3ta), a method in which three-taxon statements are produced from a nonadditive binary coding of the original data, has been proposed as a model-free way of assessing monophyly of groups, utilizing the taxic concept of homology. In fact the taxic concept amounts to a model, and, further, one that seems to conflict directly with evolution. M3ta is a type of grouping by all similarities and, like all such methods, would require a clock assumption if the tree were to be interpreted phylogenetically. Groupings based on this method, consequently, are phenetic, and they have little to do with monophyly. It has been proposed to define phylogenetic systematics in terms of grouping only by presences. While popular among advocates of 3ta, such definitions are completely inadequate, both because absences may be apomorphic and because phenetic methods can disagree with phylogenetic ones even when no absences are involved. © 1999 The Willi Hennig Society

SYMMETRY

Carine and Scotland (1999) introduce a new method, the modified three-taxon approach (m3ta). To demonstrate its benefits, they analyze the matrix of Fig. 1, which Kluge (1994, his Table 2) used to illustrate a weakness in Nelson and Platnick's (1991) original 3ta (N/P 3ta). Unlike parsimony (Fig. 1A), N/P 3ta fails to recover groups whose synapomorphies are reversals. The N/P 3ta tree (Fig. 1B) lacks groups (I(J K)), which are distinguished by the 0 or c states of I–K in the matrix (as throughout, the outgroup O is supposed to have plesiomorphic states for the characters shown). In contrast, the consensus of m3ta trees (Fig. 1C; cf. their Fig. 8) does have (I(J K)), and Carine and Scotland consider the problem solved:

Kluge (1994: 408) asserts that t.t.s. analysis [N/P 3ta] distorts the "... phylogenetic informativeness of evolutionary reversals..." However, both the modified t.t.s. analysis [m3ta] and the standard cladistic analysis [parsimony] recover the clades (IJK) and (JK).

Their conclusion is somewhat hasty. In this case m3ta recovers the particular groups in question, but the way in which it does so can lead to other difficulties. N/P 3ta cannot apply reversals because it produces three-taxon statement (3ts) A(BC) *only* if B and C, but not A, share the state initially presumed to be apomorphic. M3ta instead treats states symmetrically, producing 3ts A(BC) whenever B and C share *any* state not shared by A. While this allows reversals to be treated as other

О	000000000	0	cccccccc
А	100000000	А	gececece
В	110000000	В	ggccccccc
С	111000000	С	gggcccccc
D	111100000	D	ggggccccc
Е	111110000	E	gggggcccc
F	111111111	F	aaaaaaaaa
G	111111111	G	aaaaaaaaa
Н	1111111111	Н	aaaaaaaaa
Ι	011111111	I	cddddddd
J	001111111	J	ccggggggg
Κ	001111111	K	ccggggggg



FIG. 1. Matrix from Kluge (1994, his Table 2), presented first as 0/1 characters (the original version), then as nucleotide codes for selected sites. The two versions give the same results. O is the outgroup. (A) Most parsimonious tree. (B) Consensus of N/P 3ta trees. (C) Consensus of m3ta trees. (D) Most parsimonious tree as rerooted by Carine and Scotland (1999, cf. their Fig. 7).

synapomorphies, it also causes real symplesiomorphies to be treated like synapomorphies.

The consequences of m3ta's type of symmetry can be appreciated from the matrix of Fig. 2, which models the classic debate on the classification of amniotes (for a review, see Farris, 1979). The characters have been made congruent so that the grouping is clear. Two lines, A and H, have been supplied with several autapomorphies, reflecting the divergence of birds and mammals, and to make this apparent, the most parsimonious tree (Fig. 2A) is drawn with branch lengths. In the debate, phylogeneticists (for example, Hennig, 1975) pointed out that grouping should be based on synapomorphy. This gives the most parsimonious tree (Fig. 2A) (for the relationship between synapomorphy



FIG. 2. Matrix modeling amniotes. The entries are nucleotide codes for selected sites. O is the outgroup. (A) Most parsimonious tree, shown with branch lengths. (B) Consensus of UPGMA phenograms. (C) Consensus of m3ta trees.

and parsimony, see Farris, 1983, 1986; Farris and Kluge, 1985, 1986). Pheneticists (Mayr, 1974; Sokal, 1975; Michener, 1978) insisted that grouping should instead be based on all similarities. That would cause highly divergent groups to be removed from their genealogical positions, leaving residual groups based only on symplesiomorphy. This is seen in groups (BCD) and (EFG) of Fig. 2B, which is the consensus of UPGMA phenograms. But those paraphyletic groups, and the same displacement of A and H, also occur in Fig. 2C, which is the consensus of m3ta trees. The new method is sensitive to autapomorphies and can group by symplesiomorphy. Carine and Scotland have reinvented phenetics.

The same effects are found in real cases; Wheeler et al.'s (1993) data on arthropod relationships provide an example. Nephila, placed apically in the most parsimonious tree (Fig. 3), is highly divergent, so that it is separated from its closest relatives and placed more basally in the m3ta tree (Fig. 4), thereby creating a series of paraphyletic groups (beginning with Anoplodactylus). This is much like the cases of A and H in Fig. 2. The same is seen again with divergent Drosophila, except that it drags its sistergroup Papilio with it to its new position. Further, the basal split of the tree is misplaced. In the most parsimonious tree (Fig. 3) it lies between the outgroup, mollusks (Lepidochiton, Loligo), and the remaining taxa. In the m3ta tree (Fig. 4) onychophorans (Peripatus, Peripatoides) are placed with annelids (Glycera, Haemopis, Lumbricus) and mollusks rather than with arthropods (Callinectes, etc.)! The new method simply places the basal split between the phenetically most divergent groups, regardless of the actual relationships, and again this is as seen in Fig. 2.

MODELS

Carine and Scotland do not call their method phenetic, of course; they say that they have implemented the taxic view of homology. That their approach is phenetic, however, is readily apparent from the advantages that they claim for it (italics in the original)

[Parsimony] analysis implements a transformational model of character evolution. In contrast, and as Nelson (1992: 360) noted [3ta] "*is indifferent to models of character evolution*." Nelson (1992) was referring to the original N/P 3ta, but, as has just been seen (Fig. 1), that method does not recognize reversals as apomorphies. Consequently, if N/P 3ta trees were to be interpreted as phylogenetic trees, an assumption of irreversibility—which is certainly a model—would be required. This has been pointed out repeatedly (Deleporte, 1996; De Laet and Smets, 1998; Farris, 1997; Farris and Kluge, 1998), and Carine and Scotland do not maintain otherwise. In that case, however, their claim of indifference to models can mean only that 3ta trees are not intended to be interpreted phylogenetically.

Parallel observations apply to the new method. Phylogenetic application of m3ta trees does not seem to need an irreversibility assumption, but it does require another. It is well known that trees formed by grouping according to all similarities could be interpreted as phylogenies only by recourse to a clock model, and m3ta is no exception to this, as Figs. 2 and 4 illustrate. Like N/P 3ta, then, if m3ta is indifferent to models of evolution, this can be only because m3ta trees are not to be interpreted phylogenetically, but are instead purely phenetic constructs.

Another barrier to phylogenetic interpretation of Carine and Scotland's method is found in their explanation of homology (brackets in the original)

According to Patterson (1982: 34), "The taxic approach [to the study of homology] is concerned with monophyly of groups. The transformational approach is concerned with change, which need not imply grouping." Thus, for four taxa (ABCD) in which taxa A and B share character state X and taxa C and D share character state X', two groups (AB) and (CD) are hypothesized from a taxic perspective (Fig. 1).

The taxic tree would then be ((A B) (C D)), as they make clear in their Fig. 1D, but of course there are other possibilities. State X' might be a modified (or substituted) form of X, and in that case the tree could be (A B (C D)). If the opposite occurred, the tree could be (C D (A B)). The taxic choice of ((A B) (C D)) thus rests on ruling out *a priori* the possibility that either state has replaced (changed into, been substituted for) the other. If applied to nucleotide data, then, the taxic assumption would have the paradoxical implication that substitution could not have occurred at all! Consequently, attempting phylogenetic interpretation of a tree based on the taxic assumption would lead to an immediate contradiction.

Carine and Scotland might consider this reasoning



FIG. 3. Most parsimonious tree for the data of Wheeler *et al.* (1993). *Trilobita* is omitted, since it lacks nucleotide sequence data. *Loligo* and *Lepidochiton* comprise the outgroup.

inadmissible: it involves possible substitutions, whereas substitutions do not belong in their taxic view, but instead pertain to the transformational view. This does no good, however, for then the taxic view simply amounts to ruling out substitution directly, and phylogenetic interpretation of a taxic-based tree would still rest on a contradiction. Nor should this be surprising. No one but a creationist could think it realistic to exclude transformational considerations from the process of grouping. Character patterns are the product of changes. It would be astonishing if trying to analyze those patterns, while ignoring this fact, did not lead to paradox.

Writing of "clades" and "monophyly," Carine and Scotland create the impression that theirs is a phylogenetic approach, but this is entirely misleading. Their approach is phenetic, and their taxic assumption directly contradicts evolution. At the same time, they



FIG. 4. M3ta tree for the data of Wheeler *et al.* (1993) with *Trilobita* omitted. Compare Fig. 3.

criticize parsimony for involving a model, while claiming that their own methods are indifferent to models. This is again misleading, for their taxic assumption is certainly a model, if an extraordinarily ill-conceived one. If capable of phylogenetic interpretation at all, furthermore, the various forms of 3ta would require either an assumption of irreversibility or one of rate constancy, whereas parsimony needs neither. It would appear that Carine and Scotland's rejection of parsimony has, in reality, little to do with their stated reasons. Their reason is instead that, unlike their taxic assumption, parsimony is *not* based on denying evolution.

OUTGROUPS

In Kluge's (1994) analysis of the matrix in Fig. 1, the most parsimonious tree (Fig. 1 A) was rooted with an outgroup, as of course is the usual practice among phylogeneticists. In their own parsimony analysis of that matrix, in contrast, Carine and Scotland remove the outgroup and reroot the tree (their Fig. 7; cf. our Fig. 1D). They explain

For [parsimony] analysis, which produces unrooted trees, use

of an all zero outgroup is unjustifiable as both homologues [states] are potentially informative. Therefore, in the [parsimony] analysis of these data treated as paired homologues an all zero outgroup is not included. For the purposes of [m3ta] analysis, however, an all zero outgroup is included as an operational requirement of the method, and the results of [m3ta] analyses are rooted.

It is immediately apparent that the 0 states of Kluge's (1994) outgroup cannot really be the issue, inasmuch as the nucleotide version of the matrix (Fig. 1) gives the same results as the original 0/1 version. Indeed, this argument seems intended to create confusion, since Carine and Scotland use "all zero outgroup" to name two quite different things. The one utilized in m3ta is not a group at all; it would be better called a lack-all node. It is an artificial node—X in the figures—which is always constructed so that it lacks all the states present in the terminals, and it is called "all zero" because the lack-all condition is so represented in the nonadditive binary coding ("absence/presence" coding) which is used in m3ta. In contrast, the outgroups employed in parsimony analysis are terminals. They may have 0 states, but need not, as the nucleotide matrices of Figs. 1 and 2 illustrate. Even if they do have 0 states, moreover, these are states: features that may be shared with other terminals, not the lack of any such features.

If "all zero" outgroup O (Fig. 1) were represented in nonadditive binary coding (for which see Farris *et al.*, 1970) it would become a series of 1,0 entries (cf. Carine and Scotland's Tables 2 and 4),*not* a series of 0,0 entries like "all zero" X. This shows again that Carine and Scotland's comment involves two different meanings of "all zero." Notice, however, that a nonadditive binary coding is *not* normally used as input to a parsimony program. This is because doing so can lead to nonsense *all zero* reconstructed stem species that according to the coding—have no state whatever. Ironically, it is just such nodes that are deliberately used as "outgroups" in m3ta.

The differences between lack-all X and the outgroups used in parsimony analyses can be further appreciated from the m3ta trees of Figs. 1C, 2C, and 4. In all these, the original outgroups are placed apically, whereas X is located basally, as it must always be with m3ta.¹

¹X would have no such definite placement in a phylogenetic analysis. Since its coding would correspond to a node with all missing entries, it could equally well go anywhere on a most parsimonious tree.

Comparing these trees with the corresponding most parsimonious trees (Figs. 1A, 2A, and 3) also makes it clear that, although Carine and Scotland call X an "outgroup," its function is not like that of the outgroups used with parsimony. Whereas parsimony's outgroups provide a way of grouping by synapomorphy, the role of X is instead to implement the taxic aim of grouping by all similarities. As has already been seen, the "rooting" provided by m3ta's lack-all X simply places the basal split of the tree between the phenetically most divergent taxa, regardless of the actual relationships.

All these points become particularly clear in the case of Wheeler *et al.*'s (1993) data (Figs. 3 and 4). Here there is no question of an outgroup—a terminal or suite of terminals—with all 0 states, since the data are mostly nucleotide sequences.² Nor, correspondingly, is there any question of excluding the actual outgroup, *Loligo* and *Lepidochiton*, on the grounds that they are "all zero." And with those taxa included, it is obvious that the m3ta tree (Fig. 4) is "rooted" according to divergence rather than synapomorphy.

If "all zero" cannot, then, be the reason for excluding outgroups, what is the reason? Consider again Carine and Scotland's comment

For [parsimony] analysis, which produces unrooted trees, use of an . . . outgroup is unjustifiable as both homologues [states] are potentially informative.

One would almost think that they do not know that the purpose of outgroups is to root trees, but of course they do. By informative homologues they mean states that set off groups, and the idea that each state plesiomorphic or apomorphic—sets off a group is readily recognized as the taxic assumption, which is the basis of m3ta. Carine and Scotland's reason for wishing to exclude outgroups, then, is simply that outgroup rooting of most parsimonious trees does not conform to grouping by all similarities. While expressed in novel terms, it is really an ancient argument: since pheneticists do not use outgroups, they do not want phylogeneticists to do so either.

Related thinking is the basis of Carine and Scotland's remaining comment on this example. Though rerooted,

their most parsimonious tree (their Fig. 7; cf. our Figs. 1A, and 1D) still shows one disagreement with their m3ta tree (their Fig. 8; cf. our Fig. 1C)

The two results differ only in that [m3ta] resolves a clade (FGH) which is not found in the [parsimony] analysis.

Carine and Scotland argue

From a taxic perspective, both characters 1 and 2 support the groups (FGH) and (JK).

On inspecting the matrix (Fig. 1) and the most parsimonious tree (Fig. 1A), however, it is easily seen that the states common to (FGH) are simply those of the stem species of (FGHIJK). "Clade" (FGH) is merely a paraphyletic group, based only on symplesiomorphies. The "support" for that group, then, comes only from the taxic assumption—the same assumption that would imply that substitution does not occur and that grouping should be based on all similarities.

DEMARCATION

Phylogeneticists group by synapomorphy, others by all similarities, and this distinction has been well understood since Hennig (1966) called attention to it. Yet advocates of 3ta have preferred descriptions that do not mention synapomorphy (Siebert and Williams, 1998: 340):

Platnick (1985) . . . characterized cladistics as the theory of systematics under which organisms are classified by the presence of attributes, rather than some combination of the presence and absence of attributes.

The reason for resorting to such formulations is that synapomorphy is an evolutionary concept: much like Carine and Scotland, Siebert and Williams (1998: 346) criticized parsimony for using the model of "character transformation from one state to the next." (For a discussion of their views, see Farris and Kluge, 1998.) Such positions have more weaknesses than just their motives, however. Phenetic and phylogenetic methods give very different results with data such as the matrix in Fig. 2, even though there are no absences involved. Platnick's definition is thus inadequate to distinguish phenetics from phylogenetic systematics. Worse, it can be thoroughly misleading, since it would suggest that grouping by all similarities can be "cladistics." That

²The meaning of X seems uncertain in this case; it would presumably represent a "taxon" (a rock, perhaps?) with none of the nucleotides.

can hardly have been Platnick's intent; presumably he was simply unfamiliar with phenetic methods.

Interestingly, the taxic assumption seems to have arisen from a similar lack of familiarity. Patterson (1988: 83) supposed

The impact of neutral theory and the molecular clock on systematics is to make phenetic and cladistic methods equivalent at the level of DNA. That is why the cladistic analyses in Figs 4.2 and 4.3 give the same results as phenetic analysis....

Those "cladistic" analyses consisted of Patterson's own application of the taxic assumption, though he used neither that name nor 3ta. But why did he think that the clock assumption was safe? Because (same page)

Variations in rate, whether caused by stabilizing selection, altered generation length or mutation rate, or by molecular drive, can affect comparisons between taxa only during periods of independent rather than common descent, so that they should not distort the overall picture of phylogeny.

In the example of Fig. 2A, the rapid evolution of A and H occurs just in those independent lines, but the phenogram (Fig. 2B) gives a drastically incorrect impression of the phylogeny nonetheless. Patterson plainly did not know how phenetic methods function, in which case he could hardly have understood why such methods are not phylogenetic. In these circumstances, it is not surprising that he also thought that the taxic assumption was "cladistic."

Inadequacy when absences are not involved is not the only shortcoming of Platnick's conception; his definition also breaks down when absences are involved (Hennig, 1966: 94f):

[T]he transformation a-a'-a" may also consist in the complete reduction of the organ. For example, the absence of wings in fleas is undoubtedly an apomorphous character in comparison with the presence of wings in other holometabolic insects. On the other hand, the possession of wings is an apomorphous character in comparison to their absence in the so-called "Apterygota." In general we speak only of the homology of organs, but a "character" may also be the absence of an organ.

No description worded only in terms of presence and absence can succeed in characterizing phylogenetic systematics: sometimes the presence is the apomorphy, and sometime the absence is.

Misunderstanding concerning absences is important, because loss characters are frequently useful. Hennig's (1981, 1983) applied work included many examples of using such characters as evidence on phylogenetic relationships (for reviews and discussion see Farris and Kluge, 1985, 1986, 1997). Similarly, in nucleotide studies, large deletions often turn out to be valuable characters. Struwe *et al.* (1998), for example, noted that a deletion of about 100 bases in *trn*L seems to be characteristic of the Gentianaceae. Carine and Scotland, however, although they also quote Hennig's comment on fleas, have quite a different view of absences

"The absence of a character is not a character" (Nelson, 1978: 340); absence can never provide evidence of systematic relationship.

The key phrase, it develops, is "systematic relationship." For presence/absence ("complement relation") data, Carine and Scotland recommend the original N/ P 3ta. As an example they again use Kluge's (1994) 0/ 1 matrix (Fig. 1), this time taking (as Kluge did not) the 0s as absences, so that the 0s distinguishing (I(J K)) would now be losses. Carine and Scotland devote much of their paper to explaining this example, but there is no need to discuss these details, in view of their position

From a transformational perspective these data support clades (IJK) and (JK). The result of the [parsimony] analysis (Fig. 4) [see our Fig. 1A] which recovers these clades is consistent with a transformational view of homology whereas the [N/P 3ta] result (Fig. 5) [see our Fig. 1B] is not consistent with this view because it does not recover these clades.

The data support (I(J K)) if transformations are taken into account, but "absences can never provide evidence of systematic relationship." To Carine and Scotland, then, "systematic relationship" means the kind of relationship that is found in the world where presence and absence of structures are *not* the results of transformations—of evolutionary gains and losses.

Of course Carine and Scotland make no attempt to demonstrate that such a world actually exists. What they present is instead a "way of seeing." Like some of their other arguments, this is a old one dressed in new words. Compare Platnick' s (1993: 268) version

[T]he three-taxon approach does not distort data [compared to parsimony]; it merely looks at a different aspect of it (Nelson, 1992).

As Farris (1997: 140) observed

The same could be said of phenetic clustering.

Now it has been. Carine and Scotland have not introduced a new way of seeing, but reinvented an old way of not looking.

ACKNOWLEDGMENTS

We thank W. Wheeler for providing his matrix and M. Källersjö for helpful discussions of biochemistry and geology. Preparation of this paper was partially supported by NFR Grant 10204-303 to J.S.F.

REFERENCES

- Carine, M. A., and Scotland, R. W. (1999). Taxic and transformational homology: Different ways of seeing. *Cladistics* **15**,
- De Laet, J., and Smets, E. (1998). On the three-taxon approach to parsimony analysis. *Cladistics* **14**, 363–381.
- Deleporte, P. (1996). Three-taxon statements and phylogeny reconstruction. *Cladistics* 12, 273–289.
- Farris, J. S. (1979). The information content of the phylogenetic system. *Syst. Zool.* **28**, 483–519.
- 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. Columbia Univ. Press, New York.
- Farris, J. S. (1986). On the boundaries of phylogenetic systematics. *Cladistics* 2, 14–27.
- Farris, J. S. (1997). Cycles. Cladistics 13, 131-144.
- Farris, J. S., and Kluge, A. G. (1985). Parsimony, synapomorphy, and explanatory power: A reply to Duncan. *Taxon* **34**, 130–135.
- Farris, J. S., and Kluge, A. G. (1986). Synapomorphy, parsimony, and evidence. *Taxon* **35**, 298–306.

- Farris, J. S., and Kluge, A. G. (1997). Parsimony and history. *Syst. Biol.* 46, 215–218.
- Farris, J. S., and Kluge, A. G. (1998). A/the brief history of threetaxon analysis. *Cladistics* **14**, 349–362.
- Farris, J. S., Kluge, A. G., and Eckardt, M. J. (1970). A numerical approach to phylogenetic systematics. Syst. Zool. 19: 172–189.
- Hennig, W. (1966). "Phylogenetic Systematics" Univ. of Illinois Press, Urbana, IL.
- Hennig, W. (1975). "Cladistic analysis or cladistic classification?": A reply to Ernst Mayr. Syst. Zool. 24, 244–256.
- Hennig, W. (1981). "Insect Phylogeny." Wiley, New York.
- Hennig, W. (1983). "Stammesgeschite der Chordaten." Verlag Paul Parey, Hamburg.
- Kluge, A. G. (1994). Moving targets and shell games. *Cladistics* **10**, 403–413.
- Mayr, E. (1974). Cladistic analysis or cladistic classification. Zeitschrift Zool. Syst. Evol. 12, 94–128.
- Michener, C. D. (1978). Dr. Nelson on taxonomic methods. *Syst. Zool.* **27**, 112–118.
- Nelson, G. (1992). Reply to Harvey. Cladistics 8, 355-360.
- Nelson, G., and Platnick, N. I. (1991). Three-taxon statements: A more precise use of parsimony? *Cladistics* 7, 351–366.
- Patterson, C. (1988). The impact of evolutionary theories on systematics. *In* "Prospects in Systematics" (D. L. Hawksworth, Ed.), pp. 59–91. Clarendon Press, Oxford.
- Platnick, N. I. (1993). Character optimization and weighting: Differences between the standard and three-taxon approaches to phylogenetic inference. *Cladistics* 9, 267–272.
- Siebert, D. J., and Williams, D. M. (1998). Recycled. *Cladistics* 14, 339–348.
- Sokal, R. R. (1975). Mayr on cladism—and his critics. *Syst. Zool.* **24**, 257–262.
- Struwe, L., Thiv, M., Kadereit, J. W., Pepper, A. S., Motley, T. J., White, P. J., Rova, J. H. E., Potgieter, K., and Albert, V. A. (1998). *Saccifolium* (Saccifoliaceae), an endemic of Sierra de la Neblina on the Brazilian-Venezuelan border, is related to a temperate-alpine lineage of Gentianaceae. *Harvard Papers Bot.* 3, 199–124.
- Wheeler, W. C., Cartwright, P., and Hayashi, C. Y. (1993). Arthropod phylogeny: A combined approach. *Cladistics* 9, 1–40.