Herpetologica, 58(2), 2002, 270–275

SNAKE RELATIONSHIPS REVEALED BY SLOWLY-EVOLVING PROTEINS: FURTHER ANALYSIS AND A REPLY

RICHARD HIGHTON[1], S. BLAIR HEDGES[2], CARLA ANN HASS[2], AND HERNDON G. DOWLING[3]

[1]*Department of Biology, University of Maryland, College Park, MD 20742, USA*
[2]*Department of Biology, The Pennsylvania State University, University Park, PA 16802, USA*
[3]*Rendalia Biologists, 1811 Rendalia Motorway, Talladega, AL 35160, USA*

ABSTRACT: A reanalysis of our allozyme data (Dowling et al., 1996) for four slowly-evolving loci in 215 species of snakes by Buckley et al. (2000) concluded that because of ties in genetic distances our published UPGMA tree had "little resolution, indicating that these data are highly ambiguous regarding higher-level snake phylogeny." They also concluded that "the high degree of resolution in the published phenogram is an analytical artifact." Our study was intended to obtain information on lower-level relationships for the snake species that we had available, and it provided support for some current hypotheses of snake relationships at that level. Buckley et al. (2000) reached their conclusions because in their analysis they used only strict consensus trees and did not randomize the order of their input data. By randomizing data input order and using a majority-rule consensus tree, we show that there is considerable phylogenetic signal in our data.

Key words: Allozymes; Genetic distances; Phylogeny; Serpentes; UPGMA trees

FIVE years ago, we published an allozyme study of 215 species of snakes, about 8% of living species (Dowling et al., 1996). We selected four protein loci that appeared to be evolving relatively slowly because they had the fewest alleles (25–43) among the loci that we initially surveyed. With only four loci, our primary objective was not to obtain a robust phylogeny of snakes (that would be a remarkable feat for any allozyme data set surveying 215 species). Instead, our purpose was to obtain information on lower-level relationships among the species that we had available. Because of this, we entitled our study a "preliminary survey". Although we pointed out a number of protein similarities among higher-level taxa, we concluded that the interfamily-level topology of our UPGMA tree was "not well supported". We also noted that we were unable to distinguish relationships at the species level. However, at the levels of genera, tribes, subfamilies, and families, this approach provided some valuable information. We discussed the many cases of concordance with current snake taxonomy at those levels.

Recently, Buckley et al. (2000) reanalyzed our data, constructed strict consensus trees, and concluded that there was little phylogenetic signal in our data set. Thus, they claimed that the allozyme data are "highly ambiguous regarding higher-level snake taxonomy" and that the "high degree of resolution in our published phenogram is an analytical artifact" of our failure to consider alternate trees "implied by tied distance values". We disagree with these claims and will show that there is phylogenetic information in the allozyme data set.

In their reanalysis, Buckley et al. (2000) removed 92 species because, in their words, "identical taxa ($D = 1$, $I = 0$) [they meant $D = 0$, $I = 1$] provide no additional information about the structure of the tree." Actually, adding or deleting identical taxa may affect tree topology. If identical OTUs are removed from a study, the remaining OTUs may join more basal branches in a different order. Moreover, the removal of 43% of our species resulted in the deletion of the most important results of our study, which focused primarily on lower-level relationships, not on higher-level relationships as claimed by Buckley et al. (2000). Closely related species (and perhaps closely related genera) might be expected to be similar to each other at slowly evolving loci because few substitutions that result in changes in electropho-

retically detectable proteins are expected among recently diverged species. The results confirmed our expectations. At the lowest taxonomic level, there was no within-group variation in 19 groups ranging in size from 2–30 species (identical alleles at all four loci), and many other closely-related species had identical genotypes at three of the four loci. However, all species were not clustered within their own lower-level phylogenetic group; we pointed out that at least five species were likely misplaced on the tree (Dowling et al., 1996).

Ties can be a problem with some methods of phylogenetic analysis, including the one that we used (UPGMA). Tree-building algorithms often break ties arbitrarily and therefore the order in which taxa occur in the input file of the data used to calculate the tree may influence the topology of the tree (Hedges et al., 1992; Hendy et al., 1988). Buckley et al. (2000) acknowledged the possibility of this taxon-order bias, yet they did not randomize the input order of taxa. Instead they used a taxon order that was similar to the one in our study to calculate their 9999 "equivalent" trees, the maximum number for their software.

Their method did not always detect ties, even with 9999 trees. An example is the arrangement of three genera (*Carphophis, Diadophis,* and *Farancia*) on their consensus trees. These genera clustered together in our tree in the arrangement (*Carphophis* (*Diadophis, Farancia*)) and in the same way in of all their 9999 UPGMA trees (their Fig. 2) since they do not show a polytomy for the group. In their Fig. 3, their 9999 trees all had a different topology ((*Carphophis, Diadophis*) *Farancia*) indicating the input order had been reversed (the *I*-value between the pair *Carphophis* and *Diadophis* is the same as that between *Diadophis* and *Farancia*). Using their methodology, they did not find an indication of the tie involved in either of their two sets of 9999 trees in spite of the fact that each of the two arrangements would be expected half the time. In our majority-rule consensus tree (discussed below), the tie was no longer present because we included two species of *Farancia*; thus all 50 trees had the same topology. If Buckley et al. (2000) were interested in the effect of ties on the topology of UMPGA trees based on our snake data set, randomly reordering the taxa would have clearly revealed whether or not the tree topology was influenced by the presence of ties.

The majority-rule consensus method should have been used instead of the strict consensus method. As its name implies, the strict consensus method only resolves groups that appear in all sampled trees (Nei and Kumar, 2000). Thus, if the particular group of species clustered at a node is different in only one tree out of 10,000, the strict consensus tree will reflect this disagreement, resulting in a polytomy for all involved branches. We believe that a rare occurrence should not completely negate an overwhelming agreement among the majority of the trees. For example, among the 39 species of natricines in our study, *I*-values vary from 1.0 to 0 with many ties at all values (1.0, 0.75, 0.50, 0.25, 0). The same is true for the reduced number of eight species of natricines that Buckley et al. used, except that there are no *I*-values of 1.0 since they removed all species that were identical at all four loci. By using their method of analysis, the natricines form a polytomy with most or all of the remaining groups of snakes in both of their consensus trees (70 of 73 groups in their Fig. 2 and all 31 groups in their Fig. 3). In contrast, UPGMA trees will almost always cluster all 39 species of natricines (or their reduced sample of eight species) as a monophyletic group. This is because all natricines in our study have an *I*-value to at least one other natricine ≥ 0.75, while none has an *I*-value to any non-natricine snake in the study >0.50 (all natricines differ from all species in other groups at two or more of the four loci). However, by using a strict consensus tree, if only one tree in 10,000 clusters a single species of natricine in some other group (or a single species of another group within the natricines), both groups will be placed in a polytomy, indicating no resolution of the groups.

Buckley et al. (2000) noted that many of their trees agreed well with "traditional

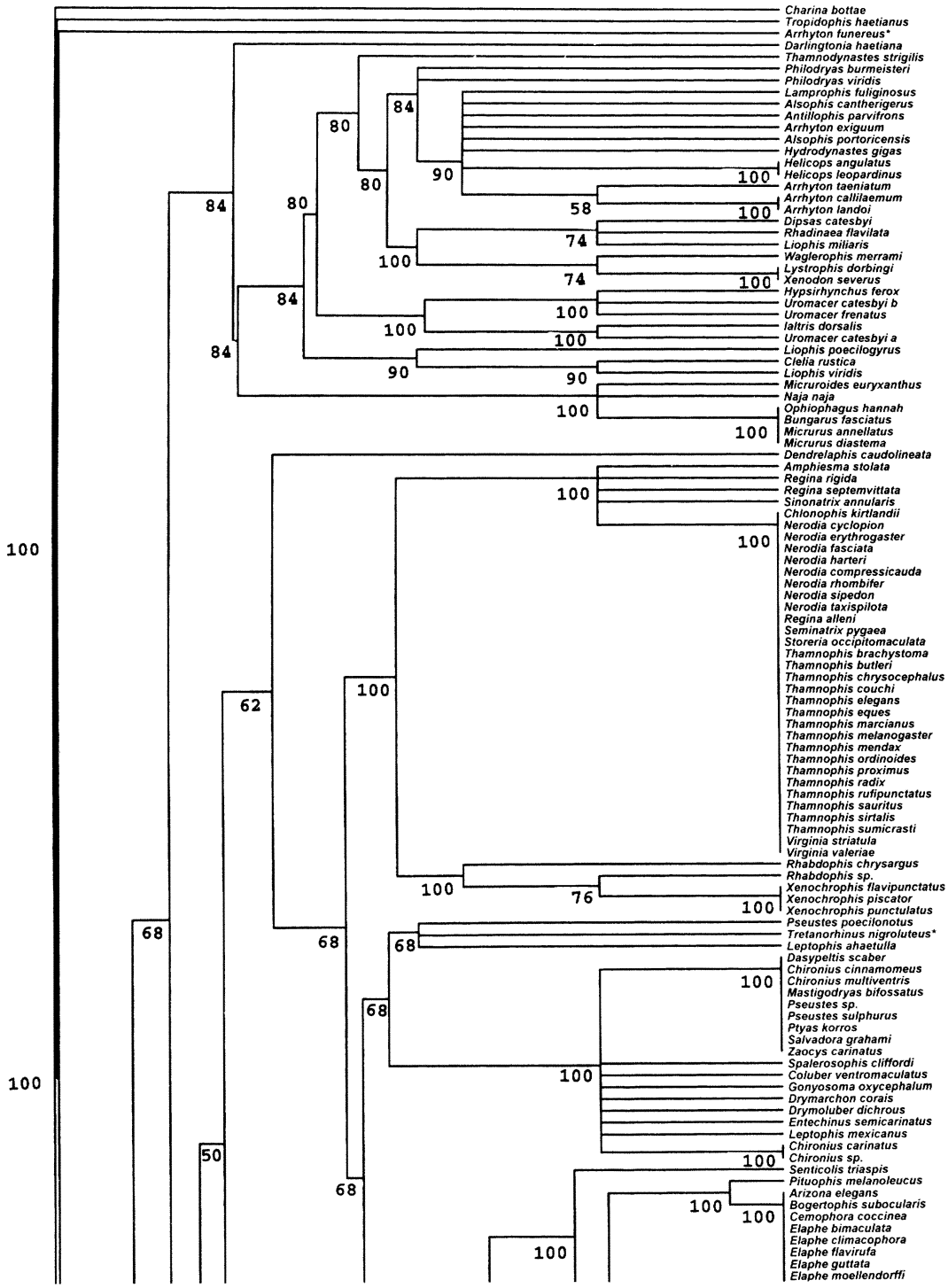

FIG. 1.—Majority rule consensus tree based on 50 UPGMA trees using Cavalli-Sforza and Edward's chord distances with randomized data input. At each node, the percentage of the 50 trees that include all the species of the group is indicated. Five species that we indicated (Dowling et al., 1996) were misplaced are indicated by asterisks. The 50 trees were calculated using Joseph Felsenstein's PHYLIP program, and the consensus tree was modified to reduce nodes occurring in <50% of the trees to polytomies.

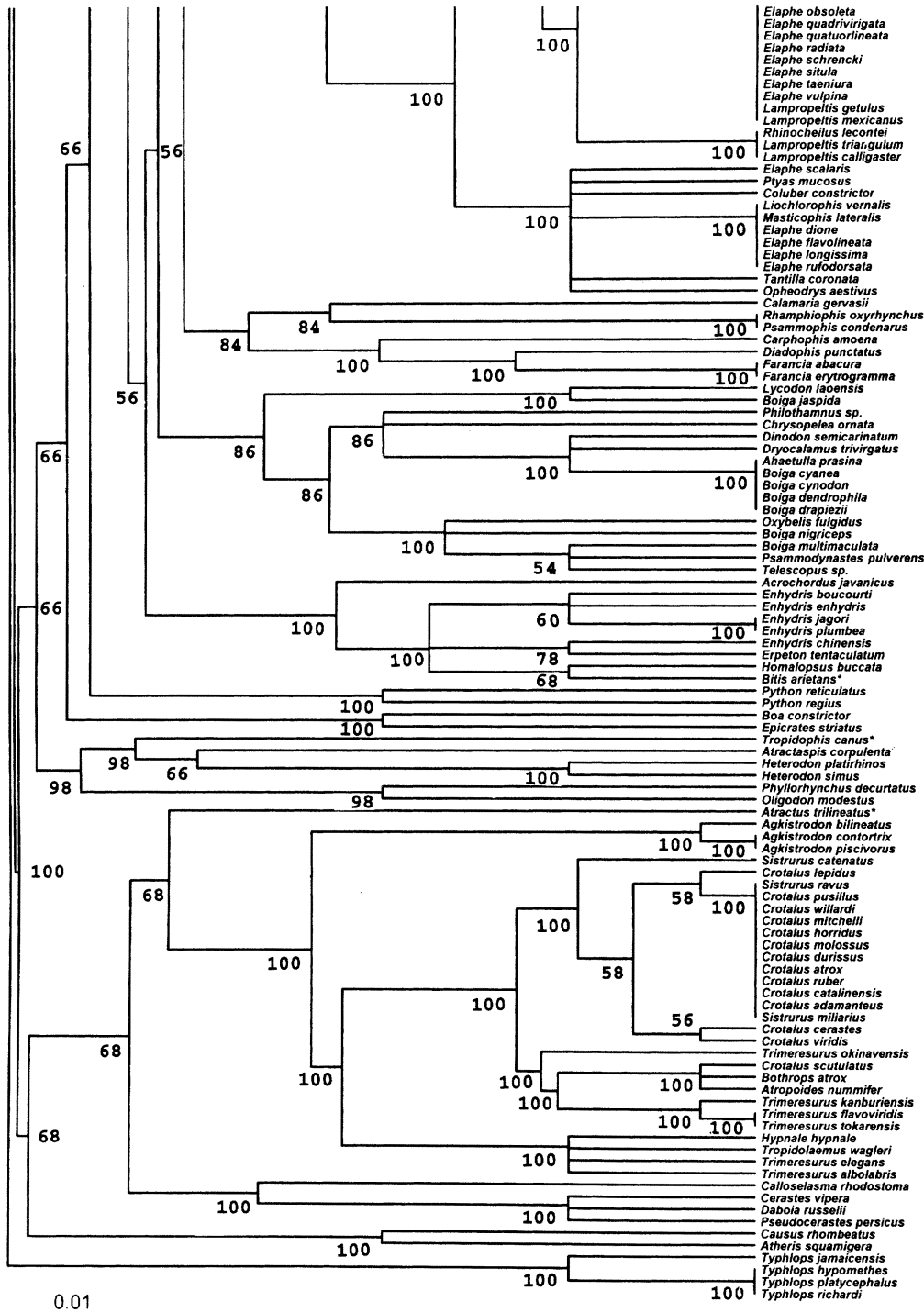

FIG. 1.—Continued.

snake taxonomy", but they still concluded that our data are not "informative about the higher level phylogeny of snakes". Two factors, (1) the pattern of genetic variation in the data, and (2) the clustering of traditional taxonomic groups together on such a high proportion of trees, make it obvious that there is phylogenetic signal present in this data set. The clustering together of all species of so many traditional groups on our UPGMA tree is not a chance event. The probability of all 39 species of natricines clustering together by chance in our single published tree is so low as to be infinitesimal, as is the clustering of many other lower-level groups on the tree. As we will show below, the same 39 species of natricines also cluster together on all 50 new trees, as do most of the other lower-level groups that we identified. A majority-rule consensus tree would have indicated the proportion of trees in which the arrangement of species at each node occurs and indicated which groups were affected by ties.

We reanalyzed our entire data set by randomizing the input of taxa 50 times and obtained an UPGMA tree for each taxon input order using Cavalli-Sforza and Edwards (1967) chord distances. A majority-rule consensus tree is shown in Fig. 1. If ties were a major problem in our data set, the resulting trees would not be in agreement with our original tree and there would be a large number of polytomies (as in the two strict consensus trees in Buckley et al., 2000). The tree we published (Dowling et al., 1996:Figs. 1–5) had 121 nodes (including the 19 nodes for the identical sets of species). The majority-rule consensus tree has 105 nodes, indicating that randomization of input order and ties in distance values have little effect on the topology of the tree for this data set. The tree in Fig. 1 is likely a better estimate of the relationships of these species than the one in Dowling et al. (1996) because rare groupings of species are eliminated and the involved nodes are collapsed into polytomies. The strict consensus trees calculated by Buckley et al. (2000) had fewer nodes, leading them to conclude that there was little resolution in our tree. The majority rule consensus tree resulting from this analysis shows that there is phylogenetic signal in the data set, in spite of the numerous ties in the distance data. Although we did not design our study to evaluate higher-level relationships and never claimed the tree to be robust, the allozyme data may provide some useful information at that level.

The UPGMA majority rule consensus tree united numerous species into clades at tribe, subfamily, or family levels, which often corresponded to traditional taxonomic groups; these were also present in our original phenogram. Some examples are (1) all 54 of the species in 26 colubrine genera of the tribe Colubrini (seen in 68% of the trees), although the group also included one misplaced dipsadine species (see discussion below); (2) all 39 species in 11 genera of the natricine tribe Natricini (100%); (3) all 16 species of nine colubrine genera of the tribes Boigini and Philothamni (86%); (4) all seven species of three homalopsine genera of the tribe Homalopsini (this group also includes an obviously misplaced viperid) (100%); (5) a group of 28 species in 18 genera of the subfamily Xenodontinae (four additional members of this subfamily are likely misplaced in this analysis—two are discussed in more detail below) (84%); (6) all six species of five genera of the family Elapidae (100%); (7) both species in one genus of the family Pythonidae (100%); (8) both species in two genera of the subfamily Boinae (100%); (9) all 30 species in 10 viperid genera of the subfamily Crotalinae (100%); (10) all four species in one genus of the family Typhlopidae (100%). Our UPGMA tree also showed that a few groups of uncertain or controversial relationships clustered with taxa with which they sometimes had been associated [e.g., *Oligodon* with *Phyllorhynchus* (98%); and *Acrochordus* with the homalopsines (100%)]. The tree also indicated the expected basal position of the family Typhlopidae (the only family to have no alleles in common with any other family), and indicated little similarity in all paired comparisons among families of snakes.

The agreement in topology among the

50 trees is high; more than half (60 of 105) of the nodes include the same grouping of species in all 50 trees. The agreement would have been even higher if it were not for the few obviously misplaced taxa. For example, we noted (Dowling et al., 1996) that two dipsadines were misplaced on our tree: *Tretanorhinus nigroluteus* clustered with the colubrines, and *Atractus trilineatus* clustered with the crotalines. These two species clustered as sister species in 16 of the 50 trees in our reanalysis. Had they not been included in our study, seven additional nodes on the path from *Tretanorhinus nigroluteus* to *Atractus trilineatus* in Fig. 1 would have been in 100% agreement on all 50 trees (instead of 68%), and another seven nodes on that same path would have much higher percentages of agreement. We obtained only 50 trees because we had to randomize the input order separately for each tree since no available computer program can do this for 215 taxa. The accuracy of the results should not be greatly affected because of using only 50 trees. For example, the expected agreement in the above example should be 66.67% (there is a three-way tie of *Tretanorhinus nigroluteus* to *Atractus trilineatus, Leptophis ahaetulla,* and *Pseustes poecilonotus,* but *Tretanorhinus* and *Atractus* will cluster together only when both are loaded consecutively, an event which would be expected one-third of the time).

Buckley et al. (2000) agreed that our "published phenogram exhibits general agreement with traditional hypotheses about snake relationships". However, no taxonomy of snakes has ever been widely accepted because there is considerable controversy in regard to the placement of many groups at most levels of classification. They mention that the thousands of "equivalent" trees that they generated also seem to be in agreement with traditional taxonomy, yet they claim that discrepancies among their trees "sum to a considerable loss of resolution". It is true that the strict consensus trees that they published have a large loss of resolution. If their thousands of UPGMA and parsimony trees are in general agreement with traditional hypotheses of snake relationships, then their use of strict consensus trees as a method of analysis would appear to have led them to the conclusion, which we dispute, that there is little resolution in our UPGMA tree. If they had used the majority-rule consensus tree approach and randomized taxon order for each tree to address the problem of ties in genetic distances, they would have retained the information that is present in the data set.

Literature Cited

BUCKLEY, L., M. KEARNEY, AND K. DE QUEIROZ. 2000. Slowly evolving protein loci and higher-level snake phylogeny: a reanalysis. Herpetologica 56: 324–332.

CAVALLI-SFORZA, L. L., AND A. W. F. EDWARDS. 1967. Phylogenetic analysis: models and estimation procedures. Evolution 21:550–570.

DOWLING, H. G., C. A. HASS, S. B. HEDGES, AND R. HIGHTON. 1996. Snake relationships revealed by slow-evolving proteins: a preliminary survey. Journal of Zoology, London 240:1–28.

HEDGES, S. B., S. KUMAR, K. TAMURA, AND M. STONEKING. 1992. Human origins and analysis of mitochondrial DNA sequences. Science 255:737–739.

HENDY, M. D., M. D. STEEL, D. PENNY, AND I. M. HENDERSON. 1988. Pp. 355–362. *In* H. H. Bock (Ed.), Classification and Related Methods of Data Analysis. Elsevier, Amsterdam, The Netherlands.

NEI, M., AND S. KUMAR. 2000. Molecular Evolution and Phylogenetics. Oxford University Press, Oxford, U.K.

Accepted: 2 October 2001
Associate Editor: Stephen Tilley

DATE OF PUBLICATION
Herpetologica, Vol. 58, No. 1, was mailed 26 February 2002.