Journal of Nematology 20(1):102–108. 1988.

Phylogenetic Analyses in Dorylaimida Using Data from 2-D Protein Patterns[1]

V. R. Ferris and J. M. Ferris[2]

Abstract: Data from two-dimensional protein patterns for nine dorylaimid isolates were analyzed using PAUP, a computer program for inferring phylogenies under the principle of maximum parsimony. With a variety of available options, including branch swapping and rooting, essentially the same tree was obtained. When isolates of the genus *Labronema* were analyzed alone, all trees obtained had the same topology, although tree length varied considerably, depending on whether a hypothetical ancestral taxon was included.

Key words: Labronema, Eudorylaimus, Aporcelaimellus, 2-D PAGE, phylogenetic analysis, PAUP, Dorylaimida.

The use of numerical phylogenetics is widespread throughout systematics, although lively controversies exist over the theoretical assumptions and justifications for many of the procedures now in use (3,4,6,15). Systematists with protein data, e.g., isozymes, usually employ algorithms that utilize some distance clustering program in which the data analyzed consist of numbers representing coefficients of overall similarity or difference that are derived from observational data. Such algorithms are, therefore, phenetic in nature. The phylogenetic method of Hennig (10), requires discrete (character) data, rather than distance data, so that character states can be polarized by some means, either a priori or as part of the analysis (7,17). It has been argued that for some biochemical data sets phenetic analyses will give the same results as phylogenetic analyses of the Hennigian type, but the issue has not been resolved (4,6). Because the phenetic approach utilizes total similarity, genealogical relationships will emerge only if rates of evolutionary divergence are constant. Hennig's method utilizes special similarity (shared, derived, homologous attributes) and makes no explicit assumptions regarding rate constancy (4).

Received for publication 12 June 1987.

[1] Journal paper 11283 Purdue University Agricultural Experiment Station, West Lafayette, IN 47907. Supported in part by NSF Grant No. BSR-8214460.

[2] Department of Entomology, Purdue University, West Lafayette, IN 47907.

We thank Mrs. Phyllis Lockard for technical assistance and Dr. D. L. Swofford for consultation on the results of our analyses using the PAUP computer program.

As a first step to augment classical morphology with biochemical data for study of evolutionary patterns in Dorylaimida, we used two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) to obtain a broad spectrum of protein data for a group of nematode isolates representing several dorylaimid taxa (5,13). Such data may be analyzed phenetically or used as discrete character data for phylogenetic analysis. Previously, we compared these protein patterns among the isolates and taxa and presented the results of a phenetic (Jaccard) analysis of the data for six isolates of *Labronema* (5). Here we analyze the same data using the computer package PAUP (Phylogenetic Analysis Using Parsimony) written by D. L. Swofford (and available from him at the Illinois Natural History Survey, Champaign, IL). As indicated by the name, PAUP resolves conflicts in the data using the principle of parsimony (14). The package finds a tree of minimal length in a Manhattan metric (a dissimilarity measure often used in numerical cladistics, also known as the city-block metric) with no a priori restrictions on the nature of permissible character-state changes. By minimizing the number of necessary character state transformations, the number of parallelisms and reversals (i.e., homoplasies) necessary to explain the evolution of each character on the postulated phylogeny is also minimized (14). The branching diagrams of the trees generated by the PAUP program can be interpreted as Hennigian cladograms. Other approaches, besides parsimony, for

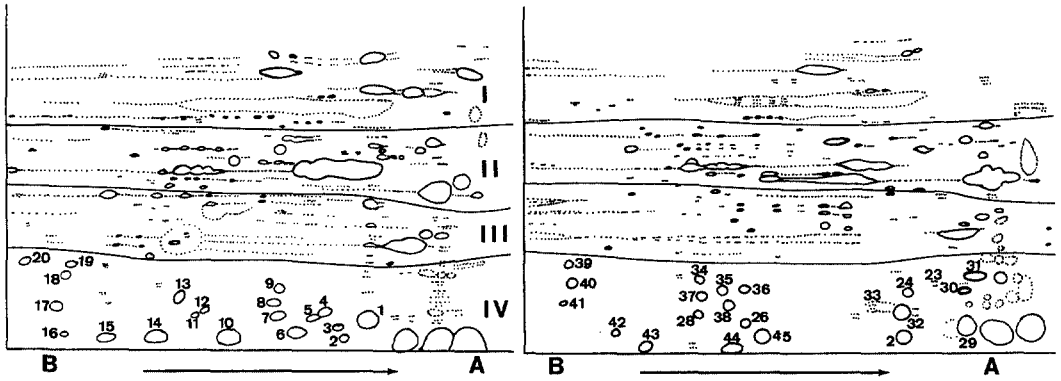

FIG. 1. Typical 2-D protein patterns for two of the isolates, Lab-IN2 (left) and Lab-FIJI (right), as discussed in text and (5). Sectors I–IV divide pattern roughly on the basis of MW, with high MW proteins at top and low MW proteins at bottom (range from ca. 70 kd to ca. 14 kd) (5). Basic proteins at left and acidic proteins at right (most in the range pH 6.7–5.0) (5). Protein spots in sector IV are numbered on sketches to illustrate technique. Inspection shows that only one protein spot in sector IV (number 2) was shared between these isolates.

dealing with conflicting data in phylogenetic analysis include probability and compatibility methods, and a considerable literature exists debating the merits of these approaches (3,4,6). We chose to use the PAUP program because of its flexibility, availability, and general sophistication (6).

MATERIALS AND METHODS

Isolates: Sources and culturing methods for the 10 dorylaimid isolates have been given previously and comparisons of their protein patterns discussed in detail (5). Three isolates of *Labronema vulvapapillatum* (Meyl) included two from Indiana (Lab-IN1 and Lab-IN2) and one from Scotland (Lab-EUR). Three nominal isolates of *L. pacificum* (Cobb) were from Florida (Lab-FL), Hawaii (Lab-HI) and Fiji (Lab-FIJI). In the present study we considered the four additional dorylaimid isolates for which we previously presented 2-D protein patterns (5) to be an outgroup to *Labronema.* These included two species that belong to the "granuliferous group" of *Eudorylaimus* Andrássy (16), one from Mauritius (Eud-MAUR) and one from Oahu, Hawaii (Eud-HI); and an isolate close to *Aporcelaimellus obscurus* Thorne and Swanger from Canada (Apor-CAN). Another isolate (Eud-KAU), from a high mountain trail on Kauai Island, Hawaii, had morphological characteristics of both *Eudorylaimus* and *Aporcelaimellus.* Voucher specimens for all isolates have been deposited in the Purdue Nematode Collection.

Proteins: The protein data were from the earlier study (5). In brief, patterns were obtained by two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) of radiolabeled proteins derived from samples comprised of equal numbers of adult females and males (when present). Proteins were located on gels by fluorography. Four to twelve patterns were obtained for each isolate, the transparent autoradiographs were overlaid and compared directly, and proteins and polypeptides with identical electrophoretic properties were assumed to be identical (1). A sketch of the typical pattern for each isolate, based on study of all the autoradiographs available, was prepared (e.g., Fig. 1). Each pattern was divided into four sectors for analysis (Fig. 1), each reproducible protein spot in each sector was numbered, and the presence or absence of that spot was determined for autoradiographs of all gels of all isolates. Only protein spots consistently showing good enough resolution to be easily recognized on multiple gels were scored, and large protein spots formed by coalescence of smaller spots were subdivided into separately numbered proteins as required by comparisons of all autoradiographs.

Analyses: Because the earlier study (5)

showed that the 2-D protein patterns for the two Indiana isolates of *L. vulvapapillatum* were identical, we treated these as one isolate (Lab-IN) in the computer analyses reported herein. All other isolates were treated as separate entities in the analyses because each had a unique array of proteins. We scored 166 proteins, but only those that were phylogenetically informative (i.e., those proteins shared by at least two, or fewer than n − 1 of the taxa) were used in the analyses. Analyses that included all nine isolates were based on 117 informative proteins. In analyses of the *Labronema* isolates only, 28 of these proteins were constant in the *Labronema* isolates and were not used by the computer program in the length calculations for the trees.

For inferring phylogenies under the principle of maximum parsimony, the PAUP program includes two basic categories of programs: viz., exact methods that guarantee that the shortest tree will be found and heuristic methods that do not guarantee optimality but generally require less computer time. The heuristic approach involves two steps: 1) the stepwise addition of OTUs (= isolates) to a developing tree or set of trees, and 2) the rearrangement of the tree(s) by "branch swapping" (14). For the heuristic analyses we used three options (SIMPLE, CLOSEST, and ROOTLESS) for determining the order in which OTUs are added to the tree. We used the GLOBAL branch-swapping algorithm, in which each possible subtree is removed from a tree during tree construction and reinserted at all other positions, with the goal of reducing tree length. We used this algorithm with MULPARS, which provides a means of discovering multiple minimum length trees when they exist. The exact method we used was the BRANCH-AND-BOUND option of PAUP, which is a refinement of Hendy and Penny's algorithm (9).

For rooting the trees we used three of the available options as follows: For the rooting option ANCESTOR, we included a "hypothetical ancestor" with all character states scored as plesiomorphic. The rooting option OUTGROUP does not require the user to make polarity decisions for the characters but lets the program make polarity decisions, based on a combination of designated ingroup and outgroup taxa. By allowing the outgroup structure to "float" during the analysis, the most parsimonious tree is not biased by erroneous polarity decisions imposed by the investigator (2,11,14). It is known that sometimes inclusion of outgroup taxa can introduce sufficient homoplasy to complicate the estimation of ingroup relationships. The LUNDBERG rooting option provides an escape from this problem by finding the shortest unrooted tree for the ingroup taxa and rooting it at the position where a hypothesized ancestor (only plesiomorphic character states) would join the tree (12,14).

Each protein was assigned a binary value (0,1) for each isolate, based on presence or absence of the protein in the pattern of the isolate. Two data sets were necessary to explore the PAUP options discussed above. In the first data set, used whenever the analysis was rooted by means of a hypothetical ancestor, the presumed plesiomorphic state was coded 0 and the apomorphic state coded 1 for each protein. In some cases presence of the protein was coded as the plesiomorphic state (0), and in other cases absence was considered to be plesiomorphic. Polarity decisions were made by the outgroup and the functional outgroup methods (8,11,17,18). These methods are based on the principle that if a character occurs in the close relatives of the group being analyzed, then the character is probably plesiomorphic, with the alternate condition apomorphic. In addition to the functional outgroup, comprised of the four isolates other than *Labronema*, we also examined autoradiographs of 2-D protein gels of additional isolates in *Eudorylaimus, Aporcelaimellus, Mesodorylaimus, Thonus, Actinolaimus,* and several mononchid genera before making polarity decisions. A second data set was used whenever rooting was by a method other than use of the hypothetical ancestor. For this data set,

absence of a protein was coded 0, presence coded 1, and the program itself determined polarity, as discussed above.

Results

The cladogram (Fig. 2) was based on the tree obtained using the branch and bound algorithm of PAUP, with the hypothetical ancestor specified. This tree took 155 steps, with an overall consistency index (c.i.) of 0.76. The numbers and kinds of protein changes in each sector of the pattern that comprised the synapomorphies for each node are mapped on the cladogram (Fig. 2). A + sign beside a number indicates additions of proteins (not previously present at a lower node), whereas a − beside a number indicates deletions of proteins that were present at a lower node. Numbers of homoplasies (i.e., changes that occurred more than once in the cladogram) are in parentheses. Other branch swapping options, including mulpars and global swap, resulted in the same tree when rooting was by hypothetical ancestor. Use of the simple addition sequence resulted in the same tree, even with no branch swapping.

All of the shortest trees obtained using data for all nine of the taxa had the topology shown in the cladogram of Figure 2. Indeed, in 20 computer runs, each of which employed some different combination of the options available, only one showed a different topology, and that tree was several steps longer than any of the trees with the topology of Figure 2. The topological difference was a linking of Apor-CAN to the *Labronema* branch, instead of to the branch with the remaining dorylaims as in the cladogram of Figure 2. This different topology was obtained when the options of closest linkage and no swapping were combined and the tree rooted using the hypothetical ancestor. Whenever branch swapping of any kind was permitted, a topology like that of Figure 2 resulted, as well as a reduction in tree length, even when the addition sequence of closest linkage was specified. These results reflect a failure in this instance of the stepwise

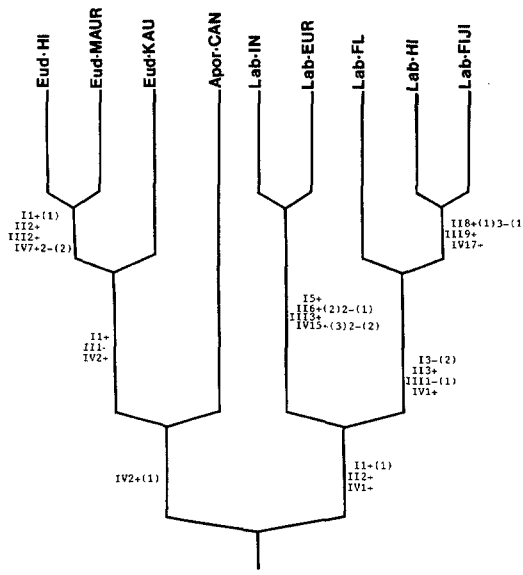

FIG. 2. Cladogram showing phylogenetic relationships of nine isolates of Dorylaimida, based on 2-D protein pattern data, with branch lengths arbitrary on diagram. Numbers of synapomorphic protein changes from each sector of the pattern for each node are listed beside the branch leading to the node. + indicates protein additions; − indicates protein deletions; numbers in parentheses indicate homoplasies.

algorithm to find the shortest tree in the absence of branch swapping.

Rooting the tree by means of the outgroup instead of the hypothetical ancestor (as discussed in Materials and Methods), together with the branch and bound branch-swapping option, produced a tree with the same topology as Figure 2, but two steps shorter (153 steps, c.i. 0.76). When the rootless addition sequence was specified and the tree rooted by means of the outgroup, the same length tree was obtained even with no branch swapping. Lundberg rooting also produced the same tree with the same statistics.

The diagrams of Figures 3 and 4 show trees obtained when the analysis included only the data for the five *Labronema* isolates. (As noted, the two Lab-IN isolates were treated as one isolate because no differences in proteins could be discerned.) These diagrams are drawn to reflect the actual branch lengths shown in the analyses, and the branch lengths are also noted on the diagrams. As in Figure 2, other

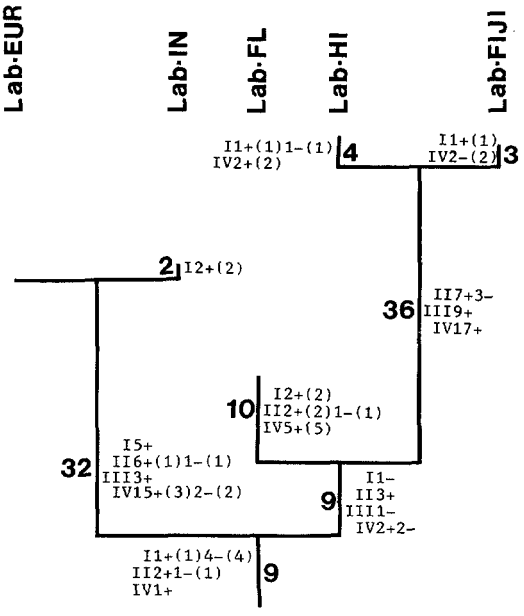

FIG. 3. Shortest tree for *Labronema* data only, when rooting was by hypothetical ancestor, with branch lengths indicated. Other notations as for Figure 2. Notations for individual isolates do not include autapomorphies.

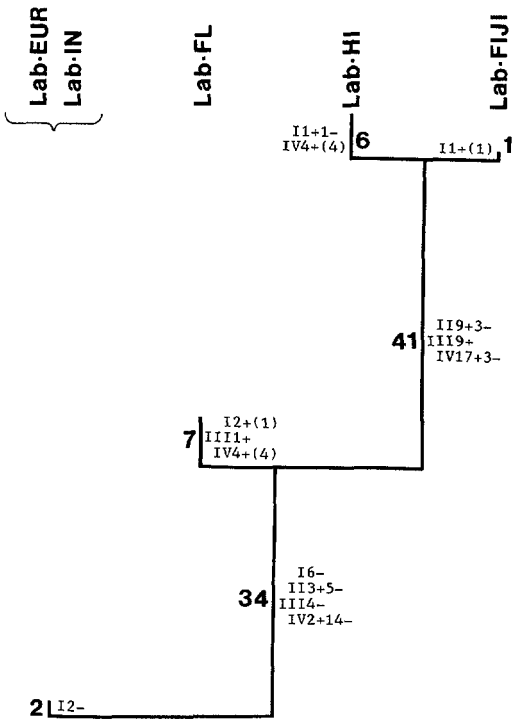

FIG. 4. Shortest tree for *Labronema* only when rooting was by outgroup, with branch lengths indicated. Other notations as in Figures 2 and 3.

numbers beside the branches leading to nodes indicate whether the synapomorphies listed involve addition or deletion of proteins. The numbers at the tips of the diagrams pertaining to the individual isolates include only those data that were also used elsewhere to derive the tree; i.e., data for autapomorphies are not listed, and the branch lengths listed do not include them.

Figure 3 shows the shortest tree obtained when the hypothetical ancestor was used for rooting. This tree is 105 steps, c.i. 0.85. A much shorter tree of 91 steps, c.i. 0.95 was obtained when the outgroup was used for rooting (Fig. 4). A tree identical to Figure 4, and with the same statistics, was also obtained using Lundberg rooting. Striking differences between the trees of Figures 3 and 4 are the relative lengths in the two trees of the branch leading to the node below Lab-IN and Lab-EUR and that leading to the rest of the taxa. In the ancestor-rooted tree, the former branch is long and the latter relatively short, whereas the situation is reversed in the outgroup-rooted tree. As is seen by the character state changes listed along the branches, in the ancestor-rooted tree (Fig. 3) the branch leading to the node below Lab-IN and Lab-EUR is supported by 29 protein additions (only four of which are homoplasies), and three protein deletions (of which all are homoplasies). In the same tree, the branch leading to the first node below the rest of the isolates is supported by five synapomorphies that comprise protein additions and four that are deletions. In Figure 4, only two synapomorphies support the node below Lab-IN and Lab-EUR, both of which are protein deletions (though not homoplasies). In the branch leading to the node below the rest of the isolates, the five synapomorphic protein additions remain but 29 additional protein deletions are also listed as synapomorphies.

DISCUSSION

The branching pattern of Figure 2 was clearly the most parsimonious arrangement of the nine isolates analyzed by means of their 2-D protein pattern data. The similarity among nearly all our trees with re-

spect to pattern, length, and c.i. indicates that our protein data set is relatively free of "noise" (i.e., excessive homoplasy) that might interfere with attempts to demonstrate relationships among the taxa. We were surprised at the similarity of results between those analyses in which the program determined polarities (based on the data of the outgroup taxa) and those in which rooting was based on our designated hypothetical ancestor.

We have drawn the cladogram of Figure 2 so that the presumed most primitive members of each of the two groups are at the center of the figure (Lab-IN and Lab-EUR for the *Labronema* isolates and Apor-CAN for the outgroup taxa), with taxa showing increasing apomorphy toward the outside of the diagram. The arrangement is concordant with known morphological data about the group, but future data of all kinds can be used to test it. Our own unpublished data for chromosome numbers in Dorylaimida are concordant, if one assumes a trend toward reduction in numbers. We have found a haploid number of six chromosomes in Lab-IN and Lab-EUR, three in Lab-FL, and one in Lab-HI and Lab-FIJI, and a haploid number of 10 in Apor-CAN and six or seven for isolates to the left of Apor-CAN in Figure 2.

The differences in branch lengths between the trees of Figures 3 and 4 result partly from differences in the two computer runs of a large group of proteins present only in Lab-IN and Lab-EUR and not in the other taxa. We interpreted the presence of these proteins as a synapomorphy for Lab-IN and Lab-EUR and coded our hypothetical ancestor to reflect this. When the *Labronema* species were analyzed alone and outgroup or Lundberg rooting was used (i.e., the trees were not rooted by means of the hypothetical ancestor), the program considered the presence of these proteins in Lab-IN and Lab-EUR to be plesiomorphic and their loss in the other taxa as apomorphic. We note, however, that the program interpreted the polarity of these proteins as we did when all of the taxa were included in the analysis. Also, we note that the coding changes did not alter tree topology in the *Labronema* analyses. Other coding differences also contributed to the shorter tree for the *Labronema* isolates obtained by outgroup rooting. In the analysis rooted by means of the hypothetical ancestor, 15 proteins changed more than once on the tree (i.e., were homoplasious); in the analysis rooted by means of the outgroup, only five proteins changed more than once.

Decisions about polarity for any given protein are necessarily difficult, primarily because the data for outgroup taxa are so severely limited. As with most coding problems, those encountered in this study will be clarified when sufficient protein data exist about additional nematode taxa to permit a well-supported hypothesis about polarity for each of the proteins in question, or when other data are so overwhelming as to leave little coding choice.

Some phylogeneticists argue that synapomorphic characters that comprise gains are stronger than those that comprise losses, because character loss is more likely to be homoplasious. This view may have some truth to it, although we have observed in 2-D protein patterns of nematodes of various taxa that apparent evolution of taxa into new forms is frequently characterized by nonrandom loss of ancestral protein constellations. In the analysis of Figure 2, the synapomorphies at each node include some gains. In some branches these are numerous and unique, as in the branch leading to Lab-HI and Lab-FIJI.

Relative to the argument that for molecular data phylogenetic relationships can be retrieved from phenetic analysis, it is interesting to compare the cladograms with our earlier Jaccard analysis of the *Labronema* protein data (5). In that analysis the phenetic similarity of Lab-FL to all other isolates was low, but the pairwise similarity coefficients with Lab-HI and Lab-FIJI were slightly lower than those with the other isolates. In the present study, however, all analyses linked Lab-FL more closely to the Pacific isolates than to the others. An important difference between the two kinds of analyses is that the Jaccard analysis was based on ancestral (plesiomorphic) as well as derived (apomorphic) similarities,

whereas the phylogenetic analyses herein described based groupings on presumed synapomorphies only.

The cladograms presented are not intended to speak to the question of taxonomic treatment for these isolates, which we will address elsewhere. They do, however, offer guidance as to probable evolutionary relationships within the group. It is our present inclination to consider Lab-CB and Lab-EUR as geographical variants of one species and to consider Lab-HI and Lab-FIJI as one species also. Lab-FL seems clearly to be a distinct species.

Literature Cited

1. Celis, J. E., and R. Bravo, editors. 1984. Two-dimensional gel electrophoresis of proteins. Orlando, FL: Academic Press.

2. Clark, C., and D. J. Curran. 1986. Outgroup analysis, homoplasy, and global parsimony: A response to Maddison, Donoghue, and Maddison. Systematic Zoology 35:422–426.

3. Felsenstein, J. 1983. Methods for inferring phylogenies: A statistical view. Pp. 315–345 *in* J. Felsenstein, ed. Numerical taxonomy. Berlin: Springer-Verlag.

4. Ferris, V. R., and J. M. Ferris. 1987. Phylogenetic concepts and methods. Pp. 346–353 *in* J. A. Veech and D. W. Dickson, eds. Vistas on nematology. Society of Nematologists.

5. Ferris, V. R., J. M. Ferris, L. L. Murdock, and J. Faghihi. 1987. Two-dimensional protein patterns in *Labronema, Aporcelaimellus* and *Eudorylaimus* (Nematoda: Dorylaimida). Journal of Nematology 19: 431–440.

6. Fink, W. L. 1986. Microcomputers and phylogenetic analysis. Science 234:1135–1139.

7. Gaffney, E. W. 1979. An introduction to the logic of phylogeny reconstruction. Pp. 79–111 *in* J. Cracraft and N. Eldredge, eds. Phylogenetic analysis and paleontology. New York: Columbia University Press.

8. Glenn, D. R., and D. R. Brooks. 1985. Phylogenetic relationships of some strongylate nematodes of primates. Proceedings of the Helminthological Society of Washington 52:227–236.

9. Hendy, M. D., and D. Penny. 1982. Branch and bound algorithms to determine minimal evolutionary trees. Mathematical Biosciences 59:277–290.

10. Hennig, W. 1966. Phylogenetic systematics (translated by D. Davis and R. Zangerl). Urbana: University of Illinois Press.

11. Maddison, W. P., M. J. Donoghue, and D. R. Maddison. 1984. Outgroup analysis and parsimony. Systematic Zoology 33:83–103.

12. Meacham, C. A. 1984. The role of hypothesized direction of characters in the estimation of evolutionary history. Taxon 33:26–38.

13. O'Farrell, P. H. 1975. High resolution two-dimensional electrophoresis of proteins. Journal of Biological Chemistry 250:4007–4021.

14. Swofford, D. L. 1984. PAUP, Phylogenetic analysis using parsimony, version 2.3, program manual. Illinois Natural History Survey, Champaign, IL.

15. Swofford, D. L., and W. P. Maddison. 1987. Reconstructing ancestral character states under Wagner parsimony. Mathematical Biosciences, in press.

16. Tjepkema, J. P., V. R. Ferris, and J. M. Ferris. 1971. Review of the genus *Aporcelaimellus* Heyns, 1965 and six species groups of the genus *Eudorylaimus* Andrássy, 1959 (Nematoda: Dorylaimida). Research Bulletin 882, Purdue University, West Lafayette, IN.

17. Wiley, E. O. 1981. Phylogenetics. New York: John Wiley and Sons.

18. Watrous, L. E., and Q. D. Wheeler. 1981. The out-group comparison method of character analysis. Systematic Zoology 30:1–11.