

# Problems with the Cladistic Use of Riboprinting

Mark E. Siddall

*Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024*

Accepted May 20, 2001

A method of character acquisition that is nearly unique to protozoological systematics is riboprinting. Briefly, the method consists of amplifying a portion of the tandemly arranged eukaryotic ribosomal genes with universal primers followed by examination of the banding patterns that result from digestion of the product with an array of restriction enzymes (Van den Bussche, 1991; Clark, 1992, 1993, 1997). The availability of universal primers that allow amplification of homologous sequences from a wide variety of taxa and the fact that the genes occur in multiple copies in the genome are advantageous and contribute to ease of amplification. Moreover, for studies concerned with many taxa, riboprinting may prove to be more economical than direct DNA sequencing and certainly will generate results in less time. However, the use of riboprinting is less straightforward than current applications might suggest and there are difficulties associated with the interpretation of restriction electromorph patterns. For the purpose of species or strain identification, these problems are minimal—either the electromorph patterns are the same or they are not. The extension of their use to phylogenetic applications (e.g., Brown and de Jonckheere, 1994; Clark *et al.*, 1995; Clark, 1997; Pernin and de Jonckheere, 1996; Clark and Diamond, 1997; Xiao and Desser, 2000) does not follow as readily because of how electromorph patterns relate to homology statements.

A character coded into a phylogenetic matrix is a homology statement. Central to the problem of accurate homology statements is the coding of presence or absence of individual bands as though they were separate and independent characters. A variety of pitfalls regarding the use of riboprinting exist and solutions are suggested here.

## DIGESTIONS

In their examination of microsporidian relationships, Pomport-Castillion *et al.* (1997) first amplified a region of the ribosomal repeat that includes SSU rDNA, ITS, and LSU rDNA. Electrophoresis of the uncut amplified products showed inequalities in size across the 12 taxa examined. Specifically, the 4 species of *Nosema* yielded shorter products than all of the other taxa, and *Agmasoma penaei* rendered a band of intermediate size. This difference in length is, of course, demonstrative of insertion/deletion events (INDELs) in the history of these taxa. Although this itself is worthy of phylogenetic consideration, it would be unwise to count a single INDEL more than once (Schaal, 1985; Dowling *et al.*, 1996). Figure 1 illustrates this problem for a hypothetical case. Even though each of two enzymes cuts in precisely the same homologous places in 2 taxa they

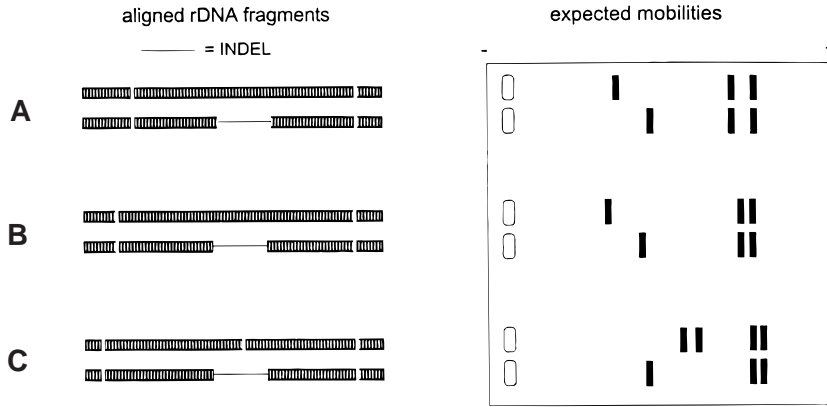


FIG. 1. When amplification products to be cut by restriction enzymes are of unequal size due to an INDEL, this single event can eventually be counted more than once. Even though there may be no difference in restriction sites for enzymes (a, b) the presence of one INDEL will cause steps to be contributed for each enzyme used. Additional steps can result from this same INDEL if an enzyme cuts within that region (c).

may yet yield different sized fragments. In Figs. 1a and 1b, two sites for each enzyme are identical for the 2 taxa. However, in both cases, the intervening INDEL (which is a single event) causes a size difference in banding patterns that would, then, be counted independently for each enzyme used. In Fig. 1c, wherein the enzyme finds a canonical site within the INDEL region of the taxon with the longer product, this event also is counted independently. For presence/absence coding of bands, this difference alone would entail an additional three steps using the presence/absence coding strategy. The sum of the number of transformations resulting from presence/absence coding for just these three enzymes is seven steps, and yet only one event has actually happened historically; the single INDEL explains all of the differences between the 2 isolates. Comparison of the banding patterns depicted by Pomport-Castillion *et al.* (1997) for *HincII*, *AluI*, *BanI*, and *MspI* digests suggests that this problem may have erroneously contributed extra steps in their analyses.

The simplest solution to the problem of unequal amplification products is to investigate alternative pairs of primers until equal-sized fragments are obtained for all study taxa prior to restriction digestion. An advantage to the use of ribosomal genes is that there are many highly conserved regions that can serve as targets for PCR primers (Hillis and Dixon, 1991). This allows for many possible permutations and many opportunities for obtaining equal-sized products. In those cases in which equal-sized products cannot be obtained, algebraic mapping of homologous fragments

(discussed more fully below) may provide a solution; otherwise it would be reasonable to admit that riboprinting cannot be used effectively and that DNA sequencing is the appropriate alternative.

Restriction enzymes recognize canonical stretches of DNA (usually palindromic sites) for digestion. However, a variety of conditions can yield bewildering results. For example, in the *AluI* digestion pattern obtained by Pomport-Castillion *et al.* (1997) (Fig. 2), for *A. panaei* (H), the two heaviest fragments are too long to sum to the size of the uncut product. Perhaps *A. panaei* exhibits heterogeneity in its ribosomal DNA (Bentzen *et al.*, 1988) and there are two mixed products in the restriction digest.

Incomplete digestion will yield more bands than corresponding restriction sites would dictate (Dowling *et al.*, 1996). Unfortunately, so too will overdigestion. Many enzymes exhibit "star formation" or cleavage at

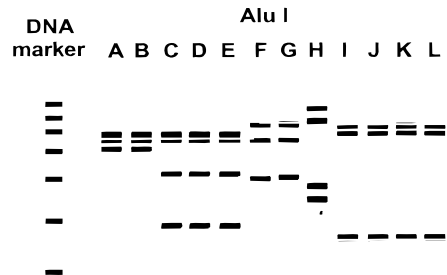


FIG. 2. The sum of electromorph fragments can exceed the total size of the uncut amplification product (as for taxon H) and lead to illogical results (redrawn from Pomport-Castillion *et al.* (1997)).



by Pomport-Castillion *et al.* (1997). Taxon A has no restriction sites and taxon C has one such site for cleavage. That is, there is a single change. However, coding the presence and absence of bands renders a minimum of three steps (loss of one band and gain of two or loss of two and gain of one, depending on the directionality of change) when logically only one transformation has occurred. This inequality would not be problematic if all events counted three steps. Yet, whereas the difference between no restriction sites and one restriction site is three steps, the difference between no sites and two sites is four steps (not six). In addition, and more difficult still, is that it is possible for two taxa to have the same restriction site, and yet have *no* electromorphs in common (Fig. 4). It would seem to be erroneous to conclude that two taxa have nothing in common when they do.

The solution is to code the restriction *site*, not the restriction *fragment*. By evaluating the actual sizes of fragments yielded by restriction digestion, it is possible, although difficult, to map the number of transformations that have occurred between any two taxa (Adams and Rothman, 1982; Templeton, 1983; DeBry and Slade, 1985; Avise, 1994; Dowling *et al.*, 1996). In the simplest case, where every taxon renders no more than two fragments, the solution is trivial, but even more complex patterns can be solved algebraically. Consider, for example, the patterns depicted by Pomport-Castillion *et al.* (1997) for the *MboI* digestion of *Spraguea lophi*

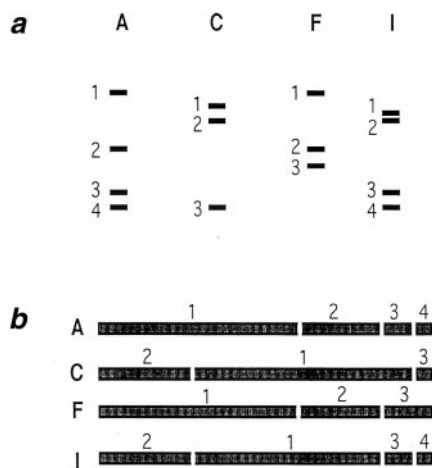


FIG. 4. Electromorph patterns (a) resulting from *MboI* digestion (redrawn after Pomport-Castillion *et al.* (1997)) can be solved algebraically (b) revealing a common restriction site for F and I even though they have no restriction fragments in common.

(A), *Glugea stephani* (C), *Microgemma ovoidea* (F), and *Nosema costelytrae* (I) reproduced here in Fig. 4a. The following homology statements obtain:

$$A_4 = C_3 = I_4,$$

$$A_1 = F_1,$$

$$A_2 = F_2,$$

$$A_3 = I_3, \text{ and}$$

$$C_2 = I_2.$$

From these we may infer

$$\text{if } A_1 = F_1, A_2 = F_2, A_4 = I_4, \text{ and } A_3 = I_3,$$

$$\text{then } F_3 = A_3 + A_4 = I_3 + I_4,$$

$$\text{if } A_4 = C_3, \text{ then } C_1 + C_2 = A_1 + A_2 + A_3,$$

$$\text{if } A_4 = I_4 \text{ and } A_3 = I_3,$$

$$\text{then } A_1 + A_2 = I_1 + I_2, \text{ and}$$

$$\text{if } C_2 = I_2 \text{ and } C_3 = I_4, \text{ then } C_1 = I_1 + I_3.$$

As a result, there is only one possible solution to the determination of restriction sites from these algebraic axioms (Fig. 4b). Note that although F and I have no fragments in common, we can infer algebraically that they do share a common restriction site. Coding of presence and absence of restriction fragments would yield the following array of binary characters and states

A 10001011

C 01010001

F 10001100

I 00110011

and would yield a consistency index of 0.78, whereas coding the restriction sites yields the matrix

A 0111

C 1001

F 0110

I 1011

and would exhibit no homoplasy whatsoever.

Although the example given proves to be tractable,

it is not likely that this always will be so or necessarily so for all taxa. An “agnostic coding” alternative is to code each enzyme as a character and each unique banding pattern as a separate state. Thus, the agnostic alternative for the four taxa considered would render the unordered matrix

A 0  
C 1  
F 2  
I 4.

The problem here, though, is the loss of any relationship among the banding patterns and the information that there is more shared between taxa than the agnostic method allows. There is a generalized algebraic shortcut to determining the *maximum* possible number of site differences (D) between any two banding patterns *i* and *j*

$$D_{ij} = F_i + F_j - 2(S_{ij}) - 2,$$

where *F* is the total number of restriction fragments for a taxon (*i* or *j*) and *S<sub>ij</sub>* is the number of shared fragments. This information can be coded by way of a Sankoff character in PAUP\* (Swofford, 1999) or SPA (Goloboff, 1996), for the *MboI* example, with taxa and states,

A a  
C b  
F c  
I d,

and with the symmetrical Sankoff cost matrix,

	a	b	c	d
a	0	3	1	2
b	3	0	4	1
c	1	4	0	5
d	2	1	5	0

The preceding corresponds reasonably to what the algebraic solution in Fig. 4b suggests but there is still a two-step overestimation of the amount of change between *Mic. ovoidea* and *N. costelytrae* (*F* and *I*, respectively).

In consideration of the preceding pitfalls that could

confound phylogeny reconstruction with respect to riboprinting it should not be surprising that the trees found by Pomport-Castillion *et al.* (1997) for their two analyses might be subject to revision in light of these alternative methods. Unlike their previous analyses that depicted arbitrarily rooted trees, I have drawn trees as unrooted networks (Figs. 5a and 6a). Use of the agnostic coding method, in which each enzyme is a character and each unique electromorph pattern is a state for that character, provided more resolution for

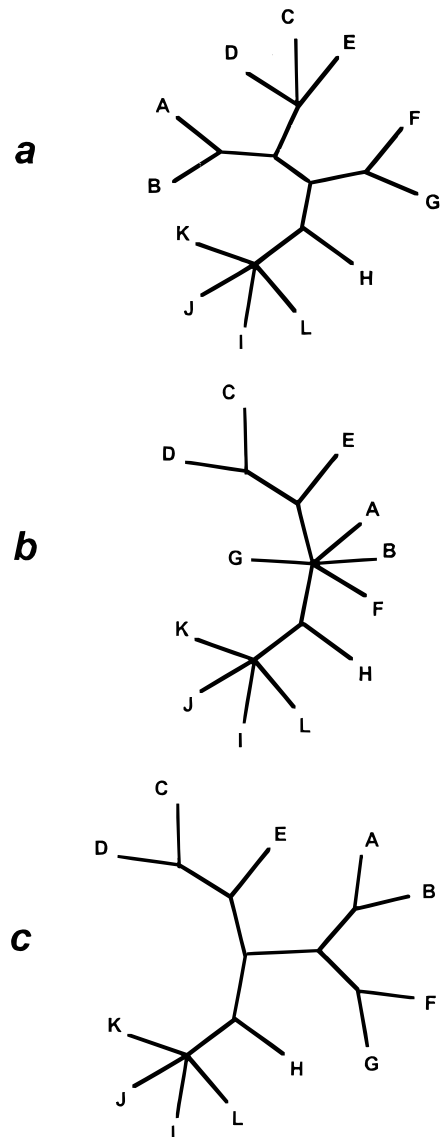


FIG. 5. Phylogenetic trees of 12 microsporidians that result from (a) presence/absence coding, (b) agnostic coding, and (c) using the Sankoff character approach.

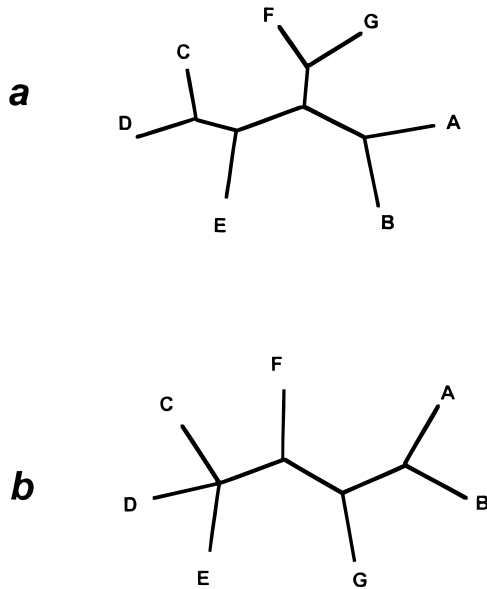


FIG. 6. Phylogenetic trees of 12 microsporidians that result from (a) presence/absence coding and (b) agnostic coding or the Sankoff character approach.

the relationships of *Glugea* species (C, D, and E) over that found by Pomport-Castillion *et al.* (1997), and de-resolved those for *Chloroscombrus* sp., *Mic. ovoidea*, *S. lophi*, and *Glugea americanus* (Fig. 5b). In contrast, except for the species of *Nosema* that were identical, the Sankoff coding method rendered a fully resolved topology (Fig. 5c). This was fully consistent with the agnostic method and which resolved different relationships for *Chloroscombrus* sp., *Mic. ovoidea*, *S. lophi*, and *G. americanus* than were found by Pomport-Castillion *et al.* (1997). In the secondary analysis, the agnostic and Sankoff methods resulted in identical solutions (Fig. 6b) that differed substantially from that found previously (Fig. 6a).

Riboprinting still is used by only a few protozoologists but its expediency and financial efficiency may yet make it appealing to others. The pitfalls identified here in no way detract from the power of riboprinting patterns for the purposes of species and strain identification. However, phylogenetic hypotheses are only as sensible as are the putative homology statements from which they are derived. Careful consideration of the enzymes used, followed by algebraically solving for restriction sites as characters instead of the presence and absence of fragments, provides the most reliable solution to otherwise confounding influences. Where

the positions of sites cannot logically be solved, it may be wise to admit to that fact and not include the results of that enzyme, or to try the agnostic and Sankoff approaches.

Ultimately it may be more reasonable to simply sequence the actual nucleotides as opposed to relying on inferential methods. By way of a simple example, Fig. 7a shows the inferred relationships of five species in the phylum Haplosporidia (*Haplosporidium nelsoni*, *Haplosporidium costale*, *Haplosporidium louisiana*, *Minchinia teredinis*, and *Urosporidium crescens*) when coding for the presence and absence of restriction sites in their known 18S rDNA sequences using *Aat*II, *Afl*III, *Apa*LI, *Ban*III, *Bcl*II, *Bgl*III, *Bsi*CI, *Bsm*I, *Bsp*MI, *Bss*HII, *Fdi*II, *Nae*I, *Nar*I, *Nco*I, *Nhe*I, *Pvu*I, *Sma*I, *Scal*, *Bst*XI, *Eae*I, *Hph*I, and *Kpn*I. In this case, because the sequences are known, the results of inferring phylogeny on the basis of restriction sites can be directly compared with the results that would be inferred from the sequenced gene (Fig. 7b). Clearly, these are not in agreement with respect to whether *Min. teredinis* is closest to *H. costale* or to *H.*

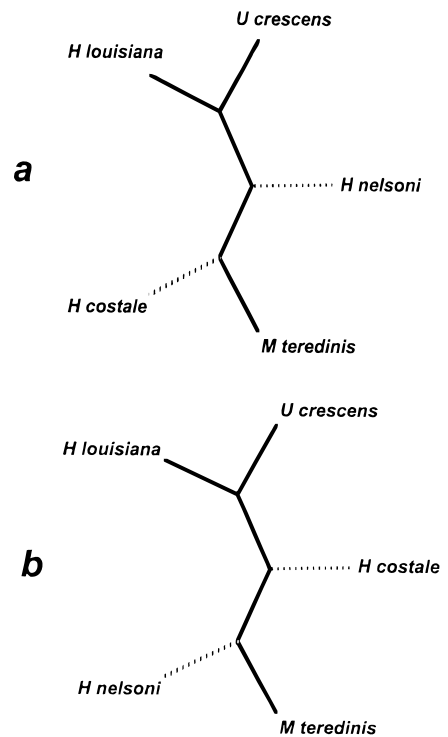


FIG. 7. Riboprinting of known 18S rDNA gene sequences from five haplosporidian species using 22 restriction enzymes (a) yields a tree that differs (dashed lines) from what is obtained with the sequenced nucleotides (b).

*nelsoni*. Insofar as the restriction sites are just an indirect assessment of sequences, it is disturbing that with only five taxa, and armed with 22 enzymes, the riboprinting method would not properly reflect the nucleotide phylogeny.

Since the riboprinting studies cited here were published, the DNA sequences for most of the organisms concerned have been completed. Comparison of the results is revealing. For example, in their study of *Entamoeba* species, Clark and Diamond (1997) could not even resolve the relationships of *E. chattoni*, *E. polecki*, and *E. coli* (Fig. 8a). Using actual sequences poses no difficulty in doing so and the results disagree substantially with those from the former method in terms of the relationships of *E. hartmanni* and *E. insolita* (Fig. 8b). Similarly, Xiao and Desser's (2000) results of riboprinting for myxozoans (Fig. 8c) are significantly different ( $P < <0.0001$  using any of the Kishino-Hasegawa, Templeton, or winning sites tests) from those found using the DNA sequence data (Fig. 8d).

Because riboprints are awkward to interpret or code

for phylogenetic analysis and because they are obviously prone to yielding spurious results that will only be overturned when someone sequences the rDNA locus, we recommend that they simply be avoided altogether in favor of the now easy and inexpensive methods of DNA sequencing.

## ACKNOWLEDGMENTS

I thank Arnold Kluge for comments on an earlier draft and the Cladistics Institute of Harbor Springs Michigan for visiting researcher status during the completion of this work. Data regarding the haplosporidians were obtained with the assistance of a grant from the National Science Foundation (BIO/DEB-9629487) in collaboration with Kim Reece and Eugene Bureson.

## REFERENCES

Adams, J., and Rothman, E. D. (1982). Estimation of phylogenetic relationships from DNA restriction patterns and selection of endonuclease cleavage sites. *Proc. Natl. Acad. Sci. USA* **79**, 3560–3564.

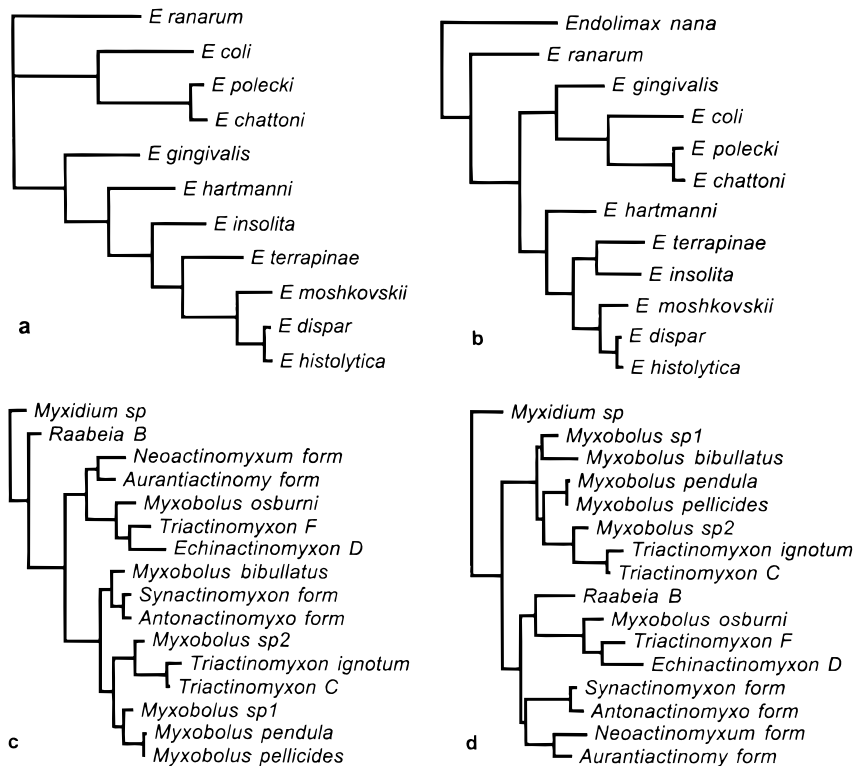


FIG. 8. Comparison of the riboprinting phylogeny (a) for *Entamoeba* species (redrawn from Clark and Diamond (1997)) with that obtained from DNA sequence data for the same locus (b), and of the riboprinting phylogeny (c) for myxozoans (redrawn from Xiao and Desser (2000)) with that obtained from DNA sequence data for this locus (d).

- Avise, J. C. (1994) "Molecular Markers, Natural History and Evolution." Chapman & Hall, New York.
- Bentzen, P. W., Legget, C., and Brown, G. G. (1988). Length and restriction site heteroplasmy in the mitochondrial DNA of American shad (*Alosa sapidissima*). *Genetics* **118**, 509–518.
- Brown, S., and de Jonckheere, J. F. (1994). Identification and phylogenetic-relationships of *Vahlkampfia* spp. (free-living amoebas) by riboprinting. *FEMS Microbiol. Lett.* **15**, 241–246.
- Clark, C. G. (1992). Riboprinting: A molecular approach to the identification and taxonomy of protozoa. In "Protocols in Protozoology" (J. J. Lee and A. T. Soldo, Eds.), Vol. 1, pp. D-4.1–D-4.4. Allen Press, Lawrence, KS.
- Clark, C. G. (1993). PCR detection of pathogenic *Entamoeba histolytica* and differentiation from other intestinal protozoa by riboprinting. In "Diagnostic Molecular Microbiology: Principles and Applications" (D. H. Persing, T. F. Smith, F. C. Tenover, and T. J. White, Eds.), pp. 468–474. ASM Press, Washington, DC.
- Clark, C. G. (1997). Riboprinting: A tool for the study of genetic diversity in microorganisms. *J. Eukaryotic Microbiol.* **44**: 2277–2283.
- Clark, C. G., and Diamond, L. S. (1997). Intraspecific variation and phylogenetic relationships in the genus *Entamoeba* as revealed by riboprinting. *J. Eukaryotic Microbiol.* **44**, 142–154.
- Clark, C. G., Martin, D. S., and Diamond, L. S. (1995). Phylogenetic relationships among anuran trypanosomes as revealed by riboprinting. *J. Eukaryotic Microbiol.* **42**, 92–96.
- DeBry, R. W., and Slade, N. A. (1985). Cladistic analysis of restriction endonuclease cleavage maps within a maximum-likelihood framework. *Syst. Zool.* **34**, 21–34.
- Dowling, T. E., Moritz, C., Palmer, J. D., and Rieseberg, L. H. (1996). Nucleic acids, III. Analysis of fragments and restriction sites. In "Molecular Systematics" (D. M. Hillis, C. Moritz, and B. K. Mable, eds.), pp. 249–320. Sinauer, Sunderland, MA. 249–320.
- Goloboff, P. (1996). SPA—Sankoff Parsimony Analysis. Instituto Miguel Lillo, Miguel Lillo 205, 4000 S. M. de Tucumon, Argentina.
- Hillis, D. M., and Dixon, M. T. (1991). DNA: Molecular evolution and phylogenetic inference. *Quart. Rev. Biol.*, **66**, 411–453.
- Hugall, A., Mortiz, C., Standton, J., and Wolstenholme, D. R. (1994). Low but strongly structured mitochondrial DNA diversity in root knot nematodes (*Meloidogyne*). *Genetics* **136**, 903–912.
- Mickevich, M. F., and Mitter, C. M. (1981). Polymorphic characters in systematics: A phylogenetic treatment of electrophoretic data. In "Advances in Cladistics I" (V. A. Funk and D. R. Brooks, Eds.), pp. 45–58, N.Y. Bot. Garden, New York.
- Mishler, B. D., Albert, V. A., Chase, M. W., Karis, P. O., and Bremer, K. (1996). Character-state weighting for DNA restriction site data: Asymmetry, ancestors and the Asteraceae. *Cladistics* **12**, 11–19.
- Murphy, R. W. (1993). The phylogenetic analysis of allozyme data: Invalidation of coding alleles by presence/absence and recommended procedures. *Biochem. Systematics Ecol.* **21**, 25–38.
- Patton, J. C., and Avise, J. C. (1983). An empirical evaluation of qualitative Hennigian analyses of protein electrophoretic data. *J. Mol. Evol.* **19**, 244–254.
- Pernin, P., and de Jonckheere, J. F. (1996). *Naegleria pussardi*, and a new *Naegleria* species phylogenetically related to the high temperature tolerant species at the molecular level. *Eur. J. Protistol.* **32**, 403–411.
- Pomport-Castillion, C., Romestand, B., and de Jonckheere, J. F. (1997). Identification and phylogenetic relationships of microsporidia by riboprinting. *J. Eukaryotic Microbiol.* **44**, 540–544.
- Schaal, B. A. (1985). Genetic variation in plant populations: From demography to DNA. In "Structure and Functioning of Plant Populations" (J. Haec and J. Woldendorp, Eds), pp. 321–342. North Holland, Amsterdam.
- Swofford, D. L. (1997). P.A.U.P\*. Phylogenetic Analysis Using Parsimony, version 4.x. Sinauer, Sunderland, MA.
- Swofford, D. L., and Olsen, G. J. (1990). Phylogeny reconstruction. In "Molecular Systematics" (D. M. Hillis and C. Moritz, Eds.), pp. 411–501. Sinauer, Sunderland, MA.
- Templeton, A. R. (1983). Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution* **37**, 221–244.
- Van den Bussche, R. A. (1991). Phylogenetic analysis of restriction site variation in the ribosomal DNA complex of New World leaf-nosed bat genera. *Syst. Zool.* **40**, 420–432.
- Weiss, J. B., Vankeulen, H., and Nash, T. E. (1992). Classification of subgroups of *Giardia lamblia* based upon ribosomal-RNA gene sequence using the polymerase chain-reaction. *Mol. Biochem. Parasitol.* **54**, 73–86.
- Xiao, C., and Desser, S. S. (2000). Molecular characterization of myxozoan parasites from Lake Sasajewun, Algonquin Park, Ontario, by riboprinting. *J. Eukaryotic Microbiol.* **47**, 85–89.