

# Long-Branch Abstractions

Mark E. Siddall\* and Michael E. Whiting†

\*Museum of Zoology, University of Michigan, Ann Arbor, Michigan 48109; and

†Department of Zoology, Brigham Young University, Provo, Utah 84602

Accepted July 11, 1998

Recent attention has been focused on the sensitivities of various tree reconstructing algorithms to sequence rate heterogeneity (long-branch attraction). Phylogenetic conclusions from two recent empirical studies have been indicted as artifacts attributable to long-branch attraction. Siddall *et al.* (1995) concluded that Myxozoa are cnidarians and sister group to *Polypodium* based on 18S rDNA and morphology. Hanelt *et al.* (1996) argued that this result is due to long-branch attraction. Whiting *et al.* (1997) concluded that the Strepsiptera are sister group to Diptera based on parsimony analysis of 18S rDNA, 28S rDNA, and morphology. Huelsenbeck (1997) argued that this result also is attributable to long-branch attraction. We demonstrate that the analyses and arguments dismissing these results as the effects of long-branch attraction are fundamentally flawed. The criteria employed by these authors were applied arbitrarily by them to the groups that they did not want, and yet using those same criteria, there is more reason to exclude other taxa besides *Polypodium* and there is more reason to disbelieve monophyly of Diptera than monophyly of Strepsiptera with Diptera. Moreover, it is asserted, long-branch attraction cannot explain the presence of nematocysts in Myxozoa and halteres in Strepsiptera. For these reasons, and in light of the demonstration that long branches cannot attract each other in their mutual absence, we conclude that the monophyly of Myxozoa + *Polypodium* and Strepsiptera + Diptera is not due to

long-branch attraction. We suggest that maximum likelihood methods are extremely sensitive to taxon and character sampling and that these data sets are demonstrative of the long-branch repulsion problem. © 1999 The Willi Hennig Society

Society

## INTRODUCTION

The conflict between cladists and likelihoodists continues unabated. Likelihoodists claim: (1) that stochastic models of DNA sequence evolution should be used to inform phylogenetic hypotheses (Felsenstein, 1973, 1981; Barry and Hartigan, 1987; Goldman, 1990) and (2) that we actually know something about how evolution works, for example, that transversions occur less frequently than transitions. Cladists claim: (1) that in the face of the infinite number of possible models (admitted to by Felsenstein, 1973), like the pheneticists before them, likelihoodists cannot avoid arbitrary choices, (2) that this itself reduces to zero the prior probability of any chosen model (Kluge and Siddall, 1997), (3) that a given set of taxa can be shown to not be in a steady-state with respect to base compositions or transition:transversion ratios such that generalizations across all taxa have no bearing on specific relationships, and (4) that because likelihood is useful for only molecular data, it disallows sampling of the full scope of

biological diversity (esp. fossils, W. Wheeler, pers. comm.) and denies the combination of all of the available information. No proponent of likelihood has ever denied the utility of such characters as "mammary glands," "vertebrae," or "amnion" as powerful evidence of the clades Mammalia, Vertebrata, and Amniota and yet neither have they answered the charge that these characters must be abandoned if we are to accept their methods (Siddall, 1997). Edwards (1992:34), though a proponent of likelihood, urged any who would apply likelihood methods to "also be influenced by the simplicity of the hypotheses, by their relevance . . . and by a multitude of subtle considerations that defy explicit statement. The scientist must be the judge of his own hypotheses, not the statistician." Popper (1983:60) was less conciliatory in his reference to a "cult of impressive technicalities," noting that "high probability is the dubious reward for saying very little, or nothing." Indeed, the mathematical calculations required for the determination of a likelihood score do seem impressive, complete with requirements for gamma distributions, Haddmard conjugations, Log-determinate corrections, and performance tuning (Swofford, *forthcoming*). Its supposed superiority, however, usually is examined with simulated datasets or with very few taxa (e.g., Huelsenbeck and Hillis, 1993; Hillis *et al.*, 1994; Gaut and Lewis, 1995; Huelsenbeck, 1995), neither of which can be of much more than passing interest to systematists who work on the relationships among real or numerous organisms.

Central to the genesis of likelihood methods and to its persistence in phylogenetics is the problem of long-branch attraction. That is, because nucleotide data are constrained to be one of four states (A, C, T, or G; insertions and deletions necessarily are ignored by likelihood methods), high rates of change, theoretically, can cause unrelated taxa to group together (Felsenstein, 1978). Only recently (Siddall, 1998) have the confounding effects of long-branch repulsion on likelihood become clear. We concern ourselves here with the problem of long-branch attraction. However, and unlike previous treatments (e.g., Huelsenbeck and Hillis, 1993; Hillis *et al.*, 1994; Gaut and Lewis, 1995; Huelsenbeck, 1995), we do so in relation to two *real* empirical problems, instead of fabricated scenarios. We assess the merits of the competing viewpoints and the consequences of the conclusions drawn by invoking long-branch attraction as an explanatory hypothesis. In one,

long-branch attraction was invoked as a rationale for excluding a taxon from analyses (Hanelt *et al.*, 1996). In the other, it was categorically claimed that long-branch attraction had been proven empirically (Huelsenbeck, 1997).

## THE MYXOZOA PROBLEM

The phylogenetic relationships of the myxozoans are of considerable importance to protozoologists, to cnidariologists, and to those studying early metazoan developmental patterns. Some myxozoans are economically important parasites of vertebrates in which they can cause debilitating pathologies like whirling disease in salmonid fishes. Thus, appropriate indications of phylogeny could lead to a better understanding of life-history patterns and of putative intervention strategies. For systematists, myxozoans are interesting because of what has happened recently regarding our understanding of their relationships. Since they were discovered, myxozoans have been the purview of protozoologists. Insofar as they were classified among the protozoa since they first were described and certainly *are* of protistan proportions, measuring only a few micrometers, this seemed appropriate (Dogeil, 1965). Recently, though, phylogenetic analyses have placed the myxozoans among the metazoa (Smothers *et al.*, 1994; Siddall *et al.*, 1995; Hanelt *et al.*, 1996), and in particular with parasitic cnidarians (Siddall *et al.*, 1995). Although some may consider this to be a testament to the power of molecular phylogenetics, as Siddall *et al.* (1995) pointed out, this hypothesis is neither surprising nor in any sense novel. Almost a century ago, Stolc (1899) remarked that these tiny parasites probably should be grouped with cnidarians because of their possession of nematocysts. This position was later argued by cnidariologists like Weill (1938) and Hadzi (1963). Electron microscopy revealed that myxozoans are multicellular with terminal cellular differentiation and that myxozoan nematocysts are identical to cnidarian nematocysts in all respects of their ontogeny as well as their mature form and function (Grasse and Lavette, 1978; Lom, 1990; Picken and Skaer, 1966; Shostak, 1993; Buss and Seilacher, 1994; Siddall *et al.*, 1995). Thus, the assertion that myxozoans are cnidarians was not in serious doubt when Siddall *et al.* (1995) provided molecular

corroboration of it. The hypothesis that myxozoans are metazoans always had been predicated on the observation of cnidarian-like characteristics. That is, Myxozoa have nematocysts. The conclusion that these structures are nematocysts is not based only on their mature morphology. The ontogeny of myxozoan nematocysts is identical to that of cnidarian nematocysts including: a preliminary external preinverted tubule and primordium, development of an electron-dense coiled thread, internalization of the tubule, and condensation of the thread into an isorhiza that has a shaft at its base and an operculum at the distal end, all of which is surrounded by a collagenous capsule (Lom, 1990; Siddall *et al.*, 1995). In addition, if myxozoans are cnidarians, and thus, metazoan, then they are apt to be closely related to the parasitic cnidarians in the Narcomedusidae. All myxozoans are parasites. All narcomedusans are parasites. Even though there is a marked diversity in nematocyst morphology among cnidarians, those of the myxozoans and the parasitic narcomedusan, *Polypodium hydriforme*, are indistinguishable in their mature atrichous isorhizal form (Megan Daly, pers. comm.). Like myxozoans, *Polypodium hydriforme* is not only a parasite, but also an intracellular parasite. Early parasitic developmental stages of narcomedusans and coelozoic myxozoans are indistinguishable (Weill, 1938; Lom, 1990). When Siddall *et al.* (1995) included *Polypodium hydriforme* in their analysis of 18S rDNA, this taxon invariably grouped with the myxozoans. Admittedly, use of fewer cnidarians for which there are complete 18S rDNA sequences rendered a paraphyletic Cnidaria; nonetheless, the myxozoans still grouped with *Polypodium*. A more judicious taxonomic sampling of cnidarians (Bridge *et al.*, 1995) rendered a monophyletic Cnidaria within which the myxozoans and *Polypodium hydriforme* were each others closest relatives (Siddall *et al.*, 1995). This was not at all surprising in light of the long-held supposition that this should in fact obtain. Thus, morphology, ultrastructure, ecology, ontogeny, ethology, and molecular data all indicate that myxozoans are most closely related to *Polypodium hydriforme*. So long as *Polypodium hydriforme* is a cnidarian (which is not in dispute), then the myxozoans necessarily are cnidarians as well.

Recently, authors of certain molecular studies have questioned this relationship. In particular, two papers appearing in the journal *Molecular Biology and Evolution*, Pawlowski *et al.* (1996) and Hanelt *et al.* (1996),

although not specifically concerned with myxozoan relationships, nevertheless discussed their phylogenetic affinities and claimed to refute the findings made by Siddall *et al.* (1995). This refutation, though, required the omission of *Polypodium* from their analyses. Lewis (pers. comm.) asserted that the grouping of *Polypodium hydriforme* with the myxozoans in the analyses of Siddall *et al.* (1995) and in their preliminary results was an artifact of long-branch length attraction. Although Hanelt *et al.* (1996) acknowledged that "Siddall *et al.* (1995) suggest that inclusion of the putative sister taxon *Polypodium* is essential to the resolution of myxozoans as cnidarians," they nonetheless excluded this taxon and concluded that their "results did not agree with the placement of the myxozoans with the phylum Cnidaria." Their rationale for exclusion of this critical taxon was their determination that the grouping must be due to "long-branch length attraction involving the *Polypodium* sequence" because "it was found to be so different from any existing 18S sequences using BLAST searches" (Hanelt *et al.*, 1995). These claims cast almost a hundred years of evidence aside as due to "long-branch length attraction."

## THE STREPSIPTERA PROBLEM

The Strepsiptera are a small (549 spp.; Kathirithamby, 1989) monophyletic group of holometabolous insects with unusual morphological, behavioral, developmental, and molecular characteristics. Associated with their lifestyle as insect endoparasitoids, Strepsiptera males have an extremely derived morphology including such characters as eyes composed of only few facets separated by cuticle or setae, mouthparts reduced and sensory or entirely absent, as well as antennae extremely flabellate and specialized (Kinzelbach, 1971). The forewings of male Strepsiptera are modified into structures which structurally are similar to, and functionally identical to, the hindwings (halteres) of Diptera (Kathirithamby, 1989; Pix *et al.*, 1993; Whiting and Wheeler, 1994). Adult female Strepsiptera are neotenic and lack mouthparts, eyes, antennae, head/thoracic segmentation, legs, and external genitalia. The extreme modification in the males and reduction in the females

is undoubtedly the reason why the phylogenetic position of the Strepsiptera has been one of the most controversial and enigmatic problems in higher level insect systematics (Kristensen, 1981, 1991). In the 19th century, Strepsiptera were aligned with such groups as Diptera, Hymenoptera, Ephemeroptera, and Siphonaptera, but by the 20th century it was assumed that Strepsiptera held a special relationship with Coleoptera, either as a group subordinate within or as the group sister to Coleoptera (reviewed in Whiting *et al.*, 1997). This supposed affinity with Coleoptera was based on a number of morphological characters which were later demonstrated to be based on incorrect descriptions of strepsipteran morphology (Kathirithamby, 1989; Pix *et al.*, 1993; Kinzelbach, 1971, 1990; Whiting and Kathirithamby, 1995; Whiting, 1998a). Currently there is only one character which is a putative homology for Coleoptera and Strepsiptera: the ability to power flight with the hindwings (postero-motorism).

Whiting and Wheeler (1994) sequenced 18S rDNA from 23 exemplar taxa representing all holometabolous orders, with three sequences from the Diptera and four from the Strepsiptera. Parsimony analysis of these data strongly supported the placement of Strepsiptera as sister group to Diptera. These authors demonstrated that the morphological data are more congruent with a placement of Strepsiptera among the Antliophoran insect orders (Diptera, Siphonaptera, Mecoptera) than with a placement with Coleoptera and suggested that this phylogenetic result may imply an unusual mode of evolution of homeotic genes in Strepsiptera.

Carmean and Crespi (1995) challenged the Strepsiptera + Diptera hypothesis with a 13-taxon analysis which included a single strepsipteran and two dipteran 18S rDNA gene sequences, but lacked any representatives of four holometabolous insect orders: Raphidioptera, Megaloptera, Trichoptera, and Lepidoptera. While these authors also found that parsimony supports Strepsiptera + Diptera, they noted that these branches were long and well-supported and thus concluded that this relationship was an artifact of parsimony's sensitivity to long-branch attraction.

Whiting *et al.* (1997) presented a more extensive study of holometabolous relationships by generating sequences from multiple representatives of all holometabolous insect orders and outgroups for 18S and 28S rDNA and by compiling an extensive morphological

matrix for the Holometabola. The molecular data included six strepsipteran and four dipteran sequences, and overall the study was a better representation of holometabolous taxonomic diversity. Whiting *et al.* (1997) found that with a variety of alignment parameters, a variety of data partitions or combinations, and a variety of weights, Strepsiptera and Diptera always formed a monophyletic group under parsimony analysis. Moreover, due to the high congruence among the multiple data sets, they concluded that the Strepsiptera + Diptera clade is currently the best supported hypothesis.

Ignoring the more comprehensive study, Huelsenbeck (1997) reexamined the same 13 sequences as Carmean and Crespi (1995), using simulation studies and maximum likelihood analyses, and again concluded that long-branch attraction is a problem with these data. Huelsenbeck (1997) concluded that Strepsiptera are sister group to Coleoptera based on his reanalysis and suggested that this result is consistent with morphological data. However, he failed to cite the more recent studies which have challenged this antiquated notion of strepsipteran morphology (reviewed in Whiting *et al.*, 1997, and in Whiting, 1998a). He concluded that it is wise to use methods that are "less sensitive to the long-branch problem in the first place, especially if application of tests proposed here indicates that long-branch attraction is a prevalent problem in nature." What the author did not justify was the extrapolation of this 13 taxon study to the more extensive study of Whiting *et al.* (1997).

## LONG-BRANCH CONFLATION AND THE MYXOZOA

Hanelt *et al.* (1996:1189) suggested that *Polypodium* should not be included because it was "so different from any existing 18S sequences using BLAST searches." However, using the BLAST search engine, *Polypodium hydriforme* is actually more like other cnidarians than myxozoans are like each other. *Polypodium* matches *Tripedalia* with a score of 895 [ $P(18) = 0.000$ ], *Hydra* with a score of 890 [ $P(16) = 0.000$ ], and *Anemonia* with a score of 850 [ $P(18) = 0.000$ ]. All of these are higher than what is found for the myxozoans which

were nonetheless left in the analysis. *Henneguya* matches *Myxidium* with a score of only 678 [ $P(16) = 0.000$ ] and *Myxobolus* with a score of 677 [ $P(15) = 1 \times 10^{-125}$ ]. *Myxidium* matched 62 plant taxa better than it matched *Myxobolus* (score of 712 [ $P(14) = 0.000$ ]) notwithstanding that the myxozoa always group together in phylogenetic analyses of this gene. Moreover, *Convoluta* does not match any of the taxa used by Hanelt *et al.* (1996) with a score better than 677 [ $P(13) = 1 \times 10^{-256}$ ] and a BLAST search using the 18S rDNA of *Strongyloides* yielded *only* fungal taxa. Thus, it would seem that if the BLAST scores and probability values are to be used as criteria for inclusion or exclusion of taxa, as specifically advocated by Hanelt *et al.* (1996), they actually had more reason to exclude other taxa than they had for the exclusion of *Polypodium*. In any case, the use of the BLAST search algorithm is a poor criterion for deciding to exclude a taxon in light of the limitations described by its providers ([http://www.ncbi.nlm.nih.gov/BLAST/blast\\_help.html](http://www.ncbi.nlm.nih.gov/BLAST/blast_help.html)):

It should be noted that the highest-scoring [sequence] ... reported in the 'High Score' column is not necessarily a member of the set of HSPs which yields the lowest P-value; the highest-scoring HSP may be excluded from this set on the basis of consistency rules governing the grouping of HSPs ... the assumption made by *blastn* that the 4 nucleotides A, C, G and T are represented at equal 25% frequencies in the database ... it should be clear that raw alignment scores are meaningless without specific knowledge of the scoring matrix that was used.

As was indicated by Hanelt *et al.* (1996), most eukaryotic 18S rDNA genes do not have equal representations of the four nucleotides. It is interesting to note that they further argued for *Polypodium*'s exclusion from analyses because of this differing nucleotide composition ( $\pi_A = 0.23$ ,  $\pi_C = 0.25$ ,  $\pi_G = 0.30$ ,  $\pi_T = 0.22$ ) and yet did not urge the exclusion of *Henneguya* which has a similar composition ( $\pi_A = 0.25$ ,  $\pi_C = 0.21$ ,  $\pi_G = 0.29$ ,  $\pi_T = 0.25$ ). Similarly, examination of genetic distances (another nonphylogenetic criterion) indicates that the mean genetic distance from *Polypodium* to the other taxa in the analysis is 0.255. This is not so different from that obtained for *Henneguya* which was left in (0.256) and is actually markedly less than the mean distance for *Convoluta* (0.279) and *Strongyloides* (0.280) also left in their analysis. In fact, the greatest distance between two taxa involves *Convoluta* (0.358 to *Dugesia*).

## LONG-BRANCH SEDUCTION AND THE STREPSIPTERA

In his reanalysis of Carmean and Crespi's (1995) limited rDNA data set of holometabolous insects, Huelsenbeck (1997) claims to have proven the first case in which long-branches have been shown to attract in nature. He claims his findings are "provocative" and have "important ramifications for phylogenetic studies" (Huelsenbeck, 1997:73). To the contrary, Huelsenbeck's (1997) findings can be shown to be the result of an uncritical approach to data analysis.

There is sufficient reason to find Huelsenbeck's (1997) premises suspect. We are instructed that the four criteria for definitively proving long-branch attraction are (1) that the branches leading to the putatively attracted groups are very long, (2) that the support for the attractors must be high, (3) that the branches in question are sufficiently long to actually attract, and (4) that some method that is (supposedly known to be) less sensitive to the long-branch attraction problem results in a tree with the branches separated. Criterion 1 is not in dispute. Criterion 2 is strange insofar as it amounts to an illogical "supposition that the clades best supported by character data are the ones we should be most suspicious of" (Whiting *et al.*, 1997:38). Following this rubric, the more evidence we have for two taxa being closely related, the more we are to disbelieve the validity of the relationship. With respect to Carmean and Crespi's (1995) data set, the branch leading to the doubted clade grouping Strepsiptera and Diptera is long and has a high bootstrap support (BS) value relative to other branches and clades for the included holometabolan taxa. We are asked to disbelieve the Strepsiptera/Diptera relationship because of the high number of synapomorphies and the high support relative to the other taxa. But the criterion is unevenly applied by Huelsenbeck. While devoting effort to the Strepsiptera/Diptera relationship, Huelsenbeck (1997) did not also note that the branch supporting Diptera (*Aedes* + *Drosophila*) is even longer and has more BS. By criterion 2, this should be less believable than the group addressed. The relative validity of the Diptera clade has bearing on criterion 3 as well.

Huelsenbeck's approach to asking whether the two branches are long enough to attract involved extracting parameters from the original tree to create a stochastic

model of change for simulated data. In those simulations, the offending branch is placed at various places on the tree. Subsequently these simulated datasets are analysed to see if the two branches attract one another irrespective of where they were placed in the model. If they do, the branches are deemed to be long enough to attract. Although not indicated by Huelsenbeck (1997), this constitutes the so-called parametric bootstrapping method of Huelsenbeck *et al.* (1996). Huelsenbeck (1997:72) found that “regardless of where Strepsiptera were [sic] placed on the simulated trees . . . branches leading to Diptera and Strepsiptera are long enough to attract each other in parsimony analysis,” and this is taken to be damning of that kind of analysis. However, these results are impossible to interpret. In the first place, only the branch representing Strepsiptera was moved around in the simulations and the dipteran clade was left in place. There are two branches that are supposed to be attracting each other. It would seem only reasonable to report the results of leaving Strepsiptera in place and moving Diptera around the tree in simulations as well. Pursuant to criticisms of criterion 2 (that the two dipterans each had branches longer than Strepsiptera), what happens when either *Aedes* or *Drosophila* alone are moved around the tree in simulations? Are these long-branched taxa also “long enough to attract” each other such that Diptera too should be abandoned?

The simulations, however, have no relationship to the original data. Their constitutive data are fictitious. That is, Huelsenbeck’s method takes parameters from the original tree and then creates new data in a Monte Carlo fashion. This will of course make for similar base compositions across all lineages in modeled trees, even though they are not in the original ( $\chi^2 = 122$ ), as well as for similar transition:transversion ratios across all lineages in the modeled trees, even though they are not in the original, and will have similar amounts of change at all sites, even though this is not true of any 18S rDNA gene nor of the original tree from which the parameters were obtained. A more reasonable approach for determining if the actual branches in the actual tree are long enough to attract is to use the actual data. This can be done by starting with the original matrix and “making a branch longer” than it is, followed by determining if this alters its location on the resulting perturbed tree. Branches can be made-long by overwriting the observed character information for

that taxon (Wenzel and Siddall, 1999). Appropriately, this is precisely the concern of long-branch attraction in the first place, that is, that the observed character states in a taxon are not due to shared history but are due to multiple overwriting substitutions having obscured informative change between speciation events, and yet this approach maintains some fidelity to the real data. When this is done repeatedly, and trees are calculated from the perturbed matrix, one can then assess whether some other branch in the tree is attracting taxa that have their character states overwritten. We did this with Carmean and Crespi’s (1995) data set (Table 1) by randomly overwriting 10 and 30% of the observed states with randomly selected nucleotides. In 1000 replicates of making the other branches in the analysis long in this fashion, it is clear that Strepsiptera is not particularly problematic. Nor for that matter is the branch leading to the dipterans. In contrast, the terminal branches leading to the individual dipterans (*Aedes* and *Drosophila*) are problematic. That is, in all of these analyses, *Aedes* or *Drosophila* showed the greatest propensity for attracting a branch made-long. This may be a result of the *Aedes* sequence being replete with

TABLE 1

Frequency (%) with Which Taxa That Have Had Their Branch Made-Long with a Specified Amount of Noise Group with Other Putative Attracting Branches in Parsimony Analyses

Branch made-long in simulations	Putatively attracting branch			
	Diptera	Strepsiptera	<i>Aedes</i>	<i>Drosophila</i>
10% Replacement with noise				
Scorpionfly	23	0	37	26
Flea	12	0	5	15
Lacewing	0	0	0	0
Antlion	0	0	0	0
Sawfly	0	0	3	0
<i>Meloe</i>	47	1	83	0
<i>Polistes</i>	33	5	38	0
<i>Tenebrio</i>	0	0	0	15
30% Replacement with noise				
Flea	59	5	63	78
Scorpionfly	83	22	87	64
Lacewing	18	3	13	29
Antlion	27	13	33	38
Sawfly	31	20	39	39
<i>Meloe</i>	90	45	94	99
<i>Polistes</i>	82	32	78	86
<i>Tenebrio</i>	60	23	22	78

sequencing artefacts (Whiting *et al.*, 1997). By Huelsenbeck's rubric, though, we have more reason to disbelieve the dipteran clade than we have for disbelieving that Strepsiptera groups with Diptera. If this is evidence that Diptera is not monophyletic it would be far more "provocative" than what Huelsenbeck (1997) detailed.

The final criterion (4 above) listed by Huelsenbeck (1997) was that some other method known to be insensitive to long-branch length attraction should be used to see if it splits the branches up. Huelsenbeck demonstrates this with the HKY85 +  $\Gamma_5$  model in a maximum likelihood procedure and shows that it does place Strepsiptera and Diptera in disparate parts of the tree. The HKY85 +  $\Gamma_5$  model is deemed (by Huelsenbeck, 1997:72) to be successful in this because it "allows for" rate heterogeneity, "allows for" nonequilibrium nucleotide frequencies, and "allows for" a transition:transversion rate bias. These are strong claims for a method that actually *imposes* a model of exactly five rates, *imposes* a particular set of nucleotide frequencies to all characters and taxa, and *imposes* a particular transition:transversion ratio in its estimations (Siddall and Kluge, 1997). In any case, these arguments too are empty. Every model used by Huelsenbeck (1997) in his maximum likelihood estimations, including Jukes-Cantor, Kimura, F81, HKY85, and HKY85 +  $\Gamma_5$  yields exactly the same topology. Clearly then the splitting of Strepsiptera and Diptera in the maximum likelihood analyses results from some phenomenon of using maximum likelihood, *not* from having chosen a model that "corrects for" something or "allows for" something else. After all, the Kimura model does not allow for different nucleotide compositions, the F81 model does not allow for different transition:transversion ratios, and the Jukes-Cantor model does not allow for much at all, yet each of these gave the same tree as the HKY85 +  $\Gamma_5$  model.

## LONG-BRANCH EXTRACTION

A criterion neglected by Huelsenbeck (1997), which should be satisfied if long-branch length attraction is to be claimed, is the trivial notion that branches can attract each other only when they are simultaneously

part of an analysis. If two branches are attracting each other, the absence of one of the branches should allow the remaining branch to place elsewhere in the pruned tree. That is, if each of the two branches individually group in precisely the same place as the other when they are allowed to stand alone in an analysis, one can hardly argue that they are attracted to this placement by the absent branch.

Although the statements made by Hanelt *et al.* (1996) regarding *Polypodium* being sufficiently different using phenetic criteria are demonstrably false, the charge that *Polypodium* groups with the myxozoa due only to long-branches is a more serious one. Using their alignment (<http://biology.unm.edu/~esloker/>), when *Polypodium* is excluded, the most parsimonious tree groups the myxozoans as sister to the Bilateria (Fig. 1a). When *Polypodium* is included, it groups with the myxozoans and these are all then sister to the Bilateria rendering a paraphyletic Cnidaria (Fig. 1b). When the myxozoans are excluded and *Polypodium* is left in, the resulting placement is the same (Fig. 1c): *Polypodium* groups next to the Bilateria rendering a paraphyletic Cnidaria in the absence of morphological data. Thus, the placement of *Polypodium* and the myxozoans is the same no matter which of these taxa is included or excluded. *Taxa cannot attract each other when they are not simultaneously part of an analysis.* This identical placement cannot be due to long-branch attraction. In fact, the emergent hypothesis of metazoan relationships does not change dramatically whether the myxozoans only are included (Fig. 1a), the myxozoans and *Polypodium* are included (Fig. 1b), or none of these are included (Fig. 1d). Examination of branch lengths again suggests that there is a more compelling reason (by this criterion) for excluding *Convoluta*, *Strongyloides*, *Meloidogyne*, *Moniliformis*, or others than there is for excluding *Polypodium*.

Applying this criterion to Carmean and Crespi's (1995) data set also is revealing (Fig. 2). Using the same procedure as Huelsenbeck (1997), by excluding all sites with gaps, all sites with ambiguities, and all sites with lowercase nucleotides, parsimony analysis results in 27 optimal trees (a consensus is shown in Fig. 2), all of which group Strepsiptera and Diptera together. However, in analyses including only Strepsiptera or including only Diptera, these taxa group in precisely the same place—with Coleoptera. In either of the taxonomically reduced data sets, there are multiple trees,

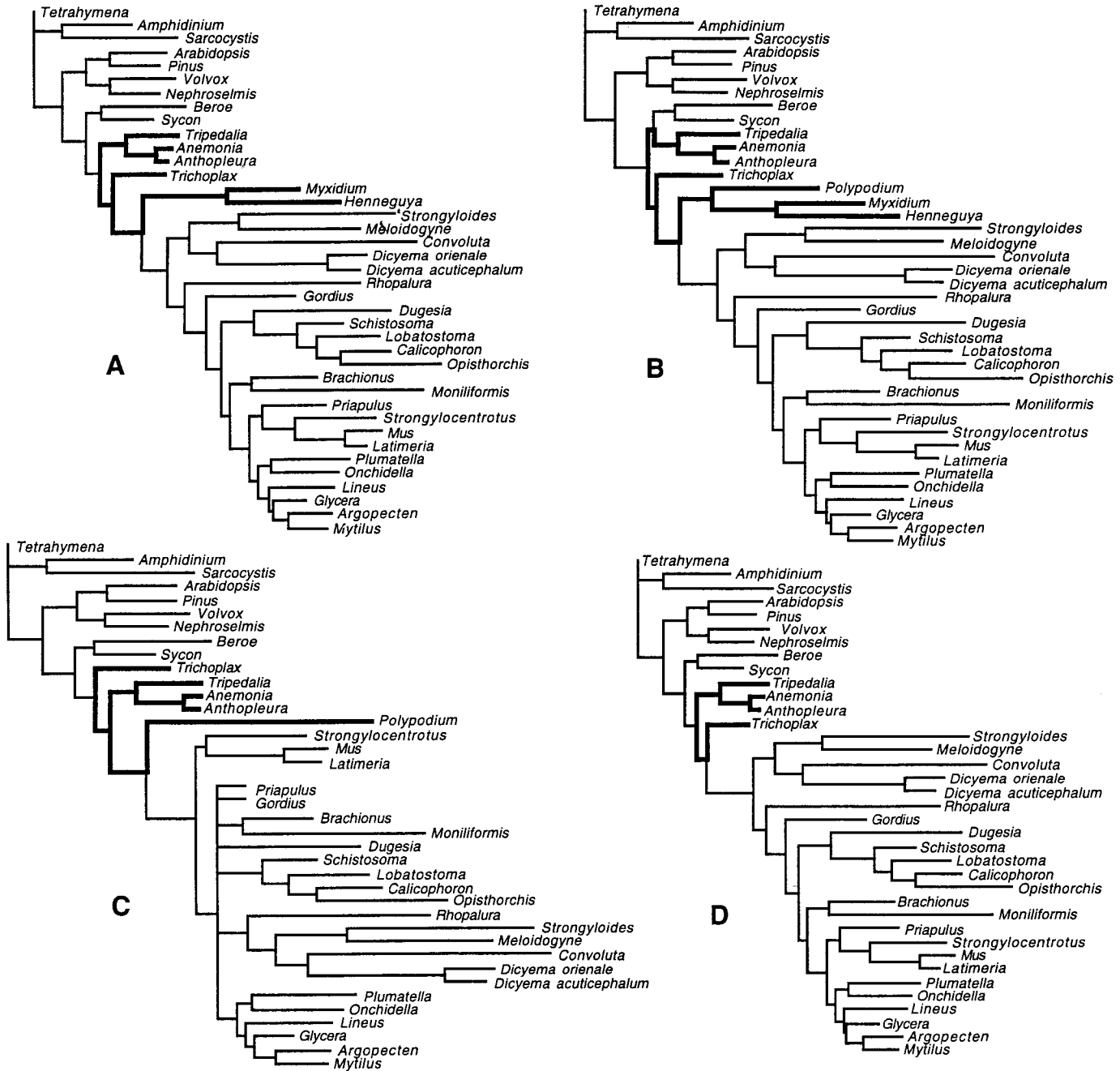


FIG. 1. Most parsimonious trees resulting when *Polypodium* is excluded (A), when *Polypodium* is included (B), when the myxozoans are excluded (C), and when all of these taxa are excluded (D). Thick branches indicate cnidarian lineages.

but Strepsiptera groups with Coleoptera when alone and Diptera groups with Coleoptera when alone. If one takes a less nihilistic approach to the data by allowing the previously excluded 67% of the data to

corroborate or refute relationships, the same result obtains. With both insect orders included, Strepsiptera and Diptera group together and with Coleoptera (Fig. 2). When Strepsiptera is excluded, Diptera groups with

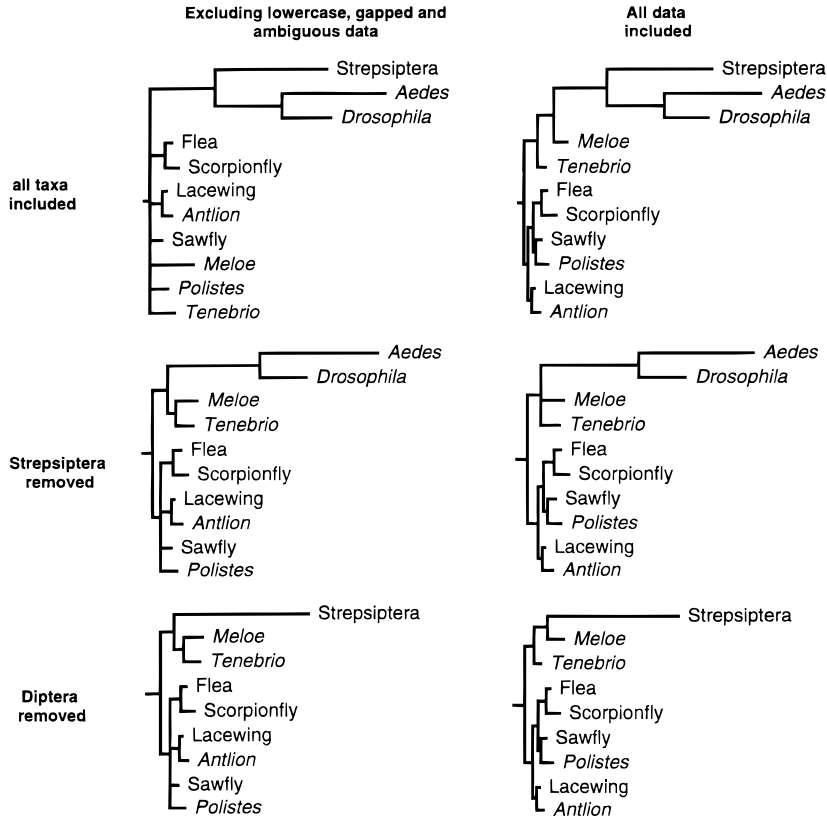


FIG. 2. Results of parsimony analyses of Carmean and Crespi's data set. Irrespective of the inclusion of taxa, Strepsiptera and Diptera group in the same place, but taxa cannot attract each other when they are not simultaneously part of the analysis.

Coleoptera. When the dipterans are excluded, Strepsiptera groups with Coleoptera. So, in fact, the grouping of Strepsiptera and Diptera together cannot be due to long-branch attraction because long-branches cannot attract when they are not simultaneously present to attract each other.

## LONG-BRANCH RETRACTION

Siddall *et al.* (1995) noted that analyses including only two or three cnidarians (Smothers *et al.*, 1994; Pawlowski *et al.*, 1996; Hanelt *et al.*, 1996) cannot be legitimately considered strong tests of the hypothesis that myxozoans are cnidarians. When Siddall *et al.* (1995) included 17 cnidarian taxa, as well as the morphological data described above, they found that the

myxozoans grouped with *Polypodium* within a monophyletic Cnidaria. Including metazoan taxa that both Siddall *et al.* (1995) and Hanelt *et al.* (1996) used for these more restricted 18S data (sans morphology) yields 15 equally parsimonious trees (length = 2183, retention index = 0.56), the strict consensus of which is shown in Fig. 3. Again, by including more cnidarians, the myxozoa fall within the group and still as sister to *Polypodium*.

Another of the authors among Hanelt *et al.* (1996) indicated that their "paper was specifically about orthonectids and their phylogenetic placement based on 18S data, not the cnidarian-*Polypodium*-myxozoan connection and all the morphological or molecular data that might pertain to that connection. We were/are not focused on that specific subject" (E. S. Loker, pers. comm.). Yet, one could argue from the results in Fig.

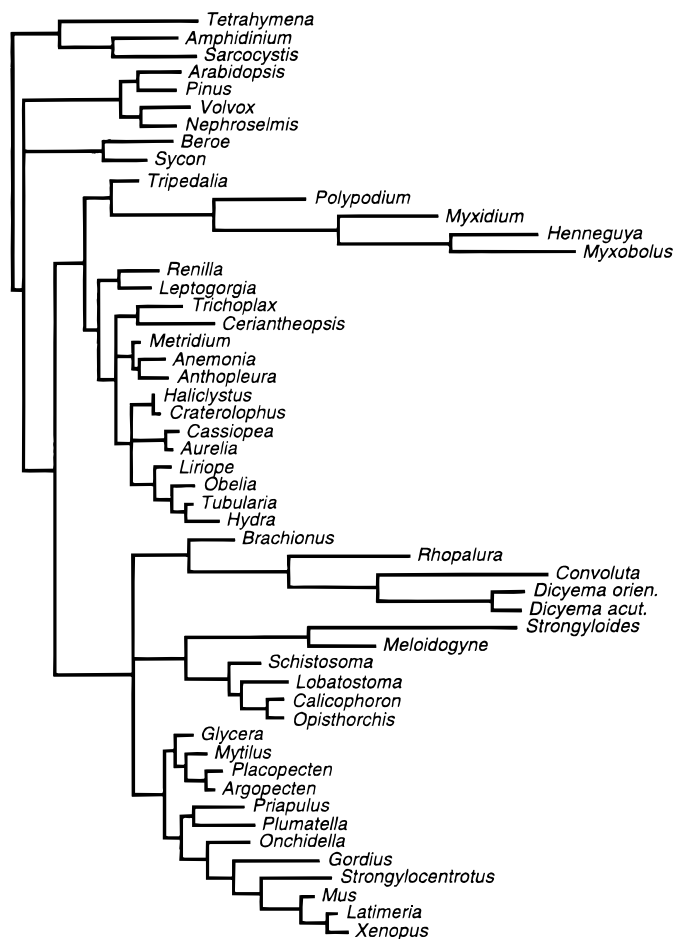


FIG. 3. Parsimony analysis of a taxonomically more inclusive molecular data set results in support for the monophyly of the myxozoans with *Polypodium hydriforme* as well as for the phylum Mesozoa.

3 that in fact they were. That is, with the broader inclusion of taxa and including *Polypodium*, the central findings of Hanelt *et al.* (1996) regarding their apparent refutation of the phylum Mesozoa are not supported. That is the orthonectid, *Rhopalura*, and the rhombozoans (*Dicyema* spp.) group together (*contra* Fig. 1a). In fact, addressing the concerns regarding long-branches, Mesozoa is supported by these data irrespective of the inclusion or exclusion of *Meloidogyne*, *Convoluta*, *Strongyloides*, the myxozoa, or *Polypodium*.

In his rush to judge maximum likelihood as a method that is “less sensitive to the long-branch problem in the first place” Huelsenbeck (1997: 74) showed that the methodology he champions is remarkably sensitive in ways not previously empirically documented. By being

only operationally concerned with the merits of parsimony and maximum likelihood, Huelsenbeck (1997) neglected any notion of character analysis. One cannot take Huelsenbeck’s findings to be indicative of “a prevalent problem in nature” if they are premised on specious homology statements. The underpinnings of any phylogenetic hypothesis are the data themselves. Although Huelsenbeck (1997:73) wondered “whether this result is robust to alignment,” it is not clear why he did not actually investigate this like Whiting *et al.* (1997; see also Wheeler, 1995) did. Rather, Huelsenbeck (1997:73) claimed that problems “with alignment, however, were minimized by excluding sites with gaps or missing data.” A cursory examination of the two alternative alignments in Table 2 should disabuse anyone of the notion that elimination of sites with gaps necessarily will compensate for illogical alignments. As a result of employing the “exclude missambig;” option in PAUP\* (Swofford, *forthcoming*), fully two-thirds of the data were left out of Huelsenbeck’s (1997) analyses. Huelsenbeck (1997) neglects to note that when all of the data are included, even likelihood groups Strepsiptera and Diptera together (Fig. 4). But maybe, after all, there are problems with the alignment, and the additional data may be “positively misleading,” causing even likelihood to converge on the answer that Huelsenbeck does not want. The whole of Huelsenbeck’s thesis, however, can be shown to rely on a single character: the 203rd character in the alignment. With only this one additional character removed from

TABLE 2

Two Alternative Alignments for the Same Data Demonstrating That Elimination of Gapped Sites Does Not Ameliorate Alignment Error

Illogical alignment	Logical alignment
-AAGGGTTT	-AAGGGTTT
AAAGGGTTT	AAAGGGTTT
AAGGATCT-	-AAGGATCT
AAGGATCT-	-AAGGATCT
AAGCATCT-	-AAGCATCT
Illogical alignment excluding “missambig”	Logical alignment excluding “missambig”
AAGGGTT	AAGGGTTT
AAGGGTT	AAGGGTTT
AGGATCT	AAGGATCT
AGGATCT	AAGGATCT
AGCATCT	AAGCATCT

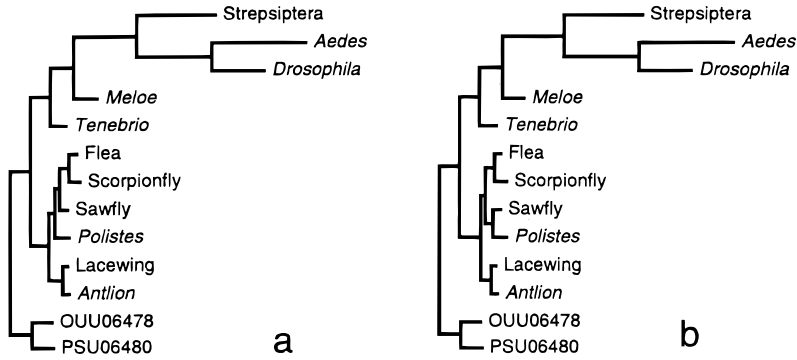


FIG. 4. Maximum likelihood analyses of the complete data of Carmean and Crespi also group Strepsiptera and Diptera together (a) using Jukes-Cantor or (b) using any of the other models or using parsimony.

Huelsenbeck's analyses, even likelihood with a HKY85 +  $\Gamma_5$  model groups Strepsiptera and Diptera together (Fig. 5). In light of this remarkable sensitivity of the likelihood method, it should surprise few that with slight realignment of the data, maximum likelihood using a Jukes-Cantor model groups Strepsiptera and Diptera together irrespective of inclusion or exclusion of characters (Fig. 6). So does the F81 model (Fig. 6). As does the Kimura model and the HKY85 model (Fig. 6). The model that "allows for" rate heterogeneity, "allows for" nonequilibrium nucleotide frequencies, and "allows for" a transition:transversion rate bias (HKY85 +  $\Gamma_5$ ) also groups Strepsiptera and Diptera together (Fig. 6).

Huelsenbeck's results are no more stable to taxonomic inclusion than they are to alignment or character inclusion. Siddall (1995) argued that if one is interested

in the stability of a hypothesis, the focus on character information (as in BS) relates to only one aspect of the data and ignores the relative importance of taxonomic sampling. The results of parsimony analyses and likelihood analyses can be compared in this way as well. Removal of any of the available taxa (save Strepsiptera itself) leaves Strepsiptera with the dipterans in parsimony analysis of Huelsenbeck's data set. Even removal of either dipteran leaves Strepsiptera with the other dipteran. Likelihood analyses, which placed Strepsiptera with Coleoptera, do not prove to be so stable. When the Jukes-Cantor or Felsenstein-81 models are employed, removal of *Aedes* returns Strepsiptera as sister to *Drosophila*, and removal of *Drosophila* places Strepsiptera with *Aedes*. Furthermore, removal of *Tenebrio* or *Polistes* or the flea places Strepsiptera with *Aedes* + *Drosophila*. Likewise, with use of the Kimura or

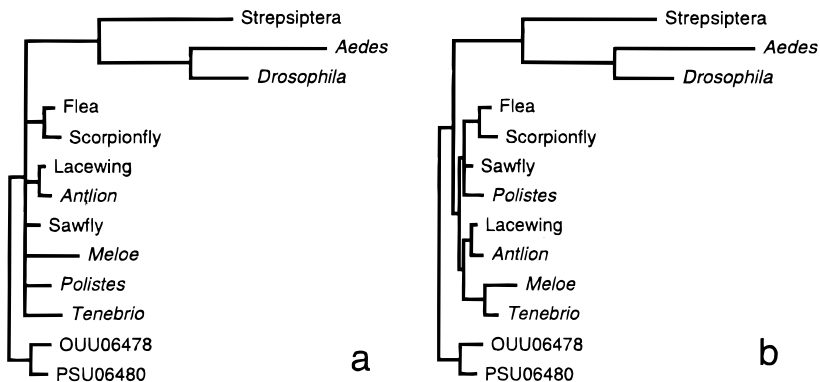


FIG. 5. The removal of one character (203) from the data set of Carmean and Crespi, in addition to those excluded by Huelsenbeck, has no effect on the results of parsimony analysis (a). Removal of this one additional character allows Strepsiptera and Diptera to group together in likelihood analyses (b).

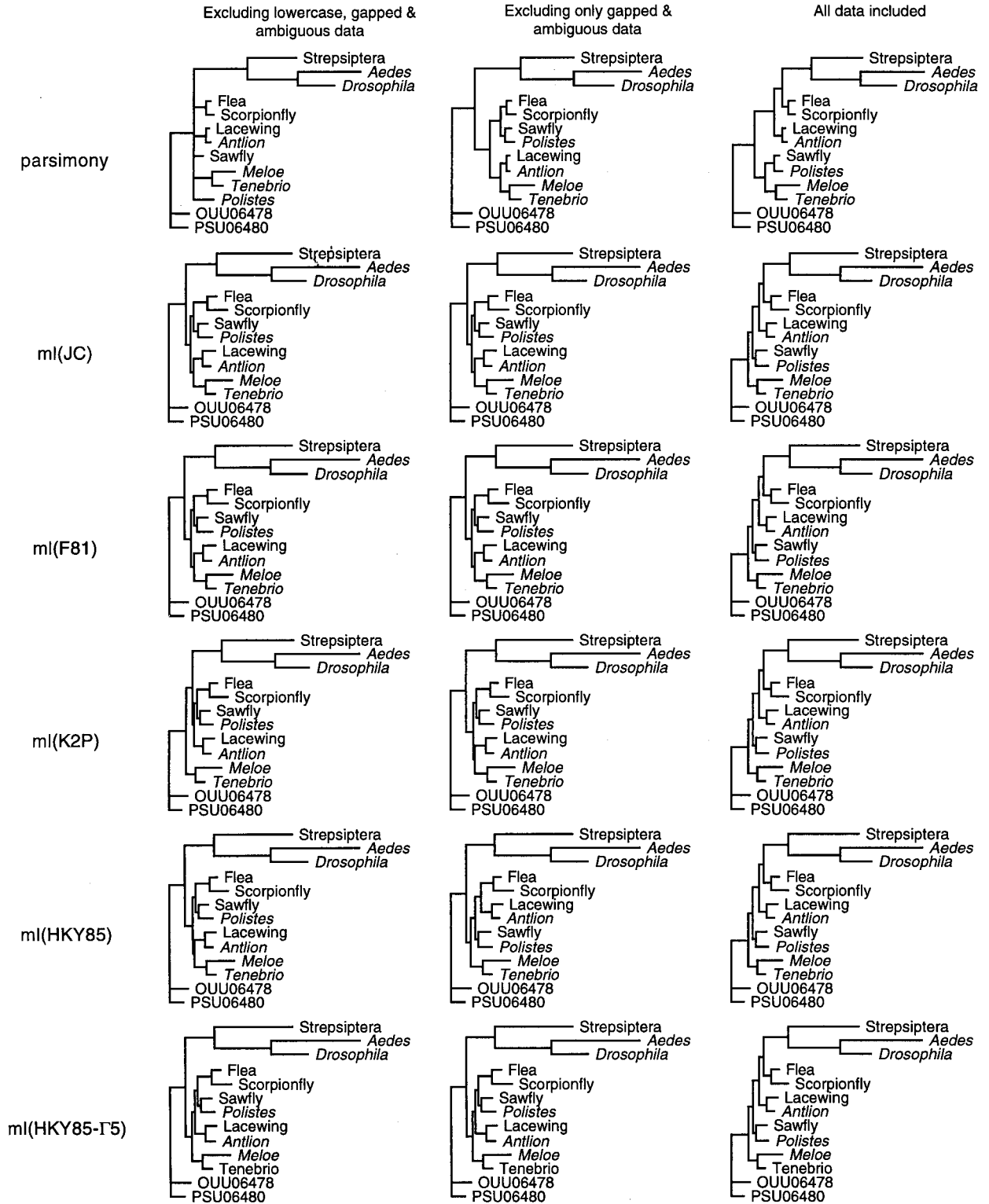


FIG. 6. Following slight realignment of the data, Strepsiptera and Diptera group together irrespective of character inclusion or of method used.

HKY85 model, removal of *Drosophila* places Strepsiptera with *Aedes*, and removal of *Tenebrio* places Strepsiptera with both dipterans.

## LONG-BRANCH REFLECTIONS

Central to the current debate is what is legitimately considered a strong test of relationships. It is argued by the probabilists that hypothesis testing consists of determining how robust a likelihood estimation is to violation of the assumptions (e.g., Swofford *et al.*, 1996). By this, it is asserted that if a tree generated according to a Jukes-Cantor model is properly estimated by a Kimura model in likelihood, the method (likelihood) must be robust. But there is a directionality to these claims that is not admitted to by Huelsenbeck (1995; nor by Huelsenbeck and Hillis, 1993, or Hillis, 1995). That is, wherein the Kimura model estimates transition:transversion ratios from the data, if those ratios are determined by a Jukes-Cantor stochastic process, the resulting Kimura model will simply approximate a Jukes-Cantor model. Finding that a more complex model properly estimates a tree generated from a simpler one is unremarkable. Having more parameters available for estimation does not mean that their estimation is material to the outcome.

Likelihood estimations require that the modeled process be true (Edwards, 1992). That is, the estimation of a tree is a conjoined hypothesis contingent on the model, not unlike the requirement for normality in the use of a *t* test in conventional statistics. By extension, if the model is not true, or can be rejected empirically, the estimation is specious and must be admitted to be so. To circumvent this difficulty, likelihood applications frequently analyze the available data with multiple models (e.g., Huelsenbeck, 1997), asserting that if all of the estimates agree, irrespective of model, then the particular choice of model is unproblematic, and the hypothesis is not contingent on the model chosen. This is a naive assertion for two reasons. First, the available models are not independent of each other. HKY85 estimates parameters for transition:transversion ratios and base compositions, Kimura estimates a parameter only for transition:transversion ratios, while F81 estimates one only for base compositions. Thus, HKY85 estimations cannot be independent of those found from K2P

and F81. But, then, what if K2P and F81 are similar? Should this not then be considered robust because they estimate different parameters? Unfortunately for the likelihoodist, transition:transversion ratios are not independent of base compositions. In Siddall and Burrenson's (1998) analysis of leech relationships, the mitochondrial sequences are extremely A and T rich ( $\pi_A = 30$ ,  $\pi_C = 17$ ,  $\pi_G = 16$ ,  $\pi_T = 38$ ). Not surprisingly, there are many more transversions than transitions (TI:TV = 0.67), the bulk of these being A:T changes. Suppose, then, that the Jukes-Cantor model also achieves a similar result as the other models, as in Huelsenbeck's analysis of the Strepsiptera. Can this then be taken to demonstrate independence of the estimate from the model chosen? In fact, no, because of the second reason: all of the models still share the remaining assertion that the nucleic acid makeup of the extant taxa is due only to a stochastic process. That is, all of the models relate to underlying processes of base change that are stochastic in nature. This denies selection, denies that one cannot have a stop codon in the middle of a gene, denies codon usage bias, denies that there are variable and conserved regions of a gene, denies that nucleic acids are obtained in a nonrandom way (i.e., sequencing), and denies historical constraint in general (Siddall and Kluge, 1997). Surely this strains credulity. The very arguments espoused by likelihoodists for incorporating models into their estimations, that we actually know something about how nucleotides change, should be reason enough to not do so. That is, from what we know about nucleotide changes we already know that there is selection, that stop codons do not occur in the middle of a gene, that there are changes in codon usage bias, that there are historical constraints on change, and that taxa are not necessarily in a steady state. In short, we know that base changes are not stochastic. Because all of the models presume that they are, it is unremarkable that they might all agree on a particular hypothesis.

A phenomenon that is common to the two problems described here, the Myxozoa and the Strepsiptera, relates directly to what we consider to be an actual strong test of a hypothesis. Hypotheses are not tested by how they stand in relation to other hypotheses but how they stand in relation to other data (Lakatos, 1970; Popper, 1983). The strongest test of a hypothesis is the acquisition and incorporation of more data. Specifically, more taxa or more characters. With respect to

the placement of the Myxozoa, the inclusion of more cnidarians only further solidifies the placement of the Myxozoa among them (Fig. 3). This was asserted from the outset (Siddall *et al.*, 1995) but was ignored by Hanelt *et al.* (1996). With respect to the Strepsiptera, Huelsenbeck's (1997) results have been shown to rely on the fewest available taxa and on a limited (proactively in this case) 18S rDNA data set. Inclusion of more taxa, morphology, and 28S rDNA continues to strengthen the hypothesis that strepsipterans are sister to the dipterans (Whiting *et al.*, 1997; Whiting, 1998b).

## LONG-BRANCH NEGATION

There is no support for the contention made by Hanelt *et al.* (1996) that myxozoans are unrelated to cnidarians. Whether one considers morphology, ecology, electron microscopy, ontogeny, ethology, or molecular data, everything points to the myxozoans and *Polypodium hydriforme* sharing a recent common ancestry. Whether one looks at morphology, 18S rDNA, or 28S rDNA, all of the available data point to Strepsiptera sharing a recent common ancestry with Diptera.

Possession of nematocysts cannot be due to long-branch attraction, any more than the identity in ontogenetic stages, or intracellular parasitism found both in the myxozoans and in the cnidarian *Polypodium hydriforme*. The use of halteres for in-flight gyroscopic balance by strepsipterans and dipterans cannot be due to long-branch attraction any more than having a ring-like ninth abdominal segment or a specialized prelabial morphology.

Like Felsenstein (1978, 1988) and others (e.g., Huelsenbeck and Hillis, 1993; Hillis *et al.*, 1994) regarding consistency, Hanelt *et al.* (1996) and Huelsenbeck (1997) fail to recognize that two long-branched taxa *might still* be each others' closest relatives.

Huelsenbeck's (1997:73) sweeping claims of a "provocative" result with "important ramifications for phylogenetic studies" is a cogent one. What has been demonstrated is not that long-branches attract in parsimony analyses—taxa cannot attract each other in their mutual absence—but that maximum likelihood analyses can be pathologically sensitive to character and taxonomic composition when sister taxa are long

enough to repel, whereas parsimony analyses are immune to this phenomenon (Siddall, 1998). To paraphrase Huelsenbeck (1997:73), the criteria outlined here should allow systematists to identify those cases in which maximum likelihood (or other methods) can be expected to have problems in the Farris Zone. However, a safer course may be to use methods that are less sensitive to the long-branch repulsion problem in the first place.

*Notes added in proof.* Huelsenbeck (*Syst. Biol.* 47, 519–537) has now reanalyzed this problem with more of the available data. That effort, which continued to ignore all morphological characters and which excluded all nonholometabolite sequences, has caused Huelsenbeck to retract his prior assertions of having demonstrated that "long-branch attraction is a real phenomenon, not just a theoretical one" (*Syst. Biol.* 46, 73). Huelsenbeck has yet to provide an explanation of the shared presence of gyroscopic halteres or other morphological synapomorphies for Diptera and Strepsiptera. Also, in a comprehensive molecular and morphological analysis of metazoa, Zravy *et al.* (*Cladistics*, 14, 249–285) have corroborated the sister group relationship of Myxozoa and Narcomedusidae as well as the monophyly of Mesozoa, contra Hanelt *et al.* (*Mol. Biol. Evol.* 13, 1187–1191).

## ACKNOWLEDGMENTS

The arguments and analyses in this paper were first detailed in a tag-team presentation by M.E.S. and M.F.W. at a symposium on likelihood at the sixteenth meeting of the Willi Hennig Society at George Washington University. We thank D. Lipscomb for not curtailing that presentation when it became clear that we would run grievously over-time. We appreciate the continued support of the Cladistics Institute of Harbor Springs Michigan where "reactionaries" are as welcome as "right and forward thinking" people. Words that loosely rhyme with "attraction" were found with the assistance of Doug Beeferman's on-line Semantic Rhyming Dictionary, which is available through the World Wide Web from the Statistical Language Modeling Group at Carnegie Mellon University (<http://www.link.cs.cmu.edu/>). We also thank David Swofford for making advance versions of PAUP\* available, without which our critiques of likelihood would be more difficult. Data relating to the mitochondrial genes of leeches were obtained through a grant from the National Science Foundation (BIO/DEB-9629487).

## REFERENCES

- Barry, D., and Hartigan, J. A. (1987). Statistical analysis of hominoid molecular evolution. *Statistical Sci.* 2, 191–210.
- Bridge, D., Cunningham, C. W., Desalle, R., and Buss, W. (1995).

- Class-level relationships in the phylum cnidaria: Molecular and morphological evidence. *Mol. Biol. Evol.* **12**, 679–689.
- Buss, L., and Seilacher, A. (1994). The phylum Vendobionta: A sister group of the Eumetazoa? *Paleobiology* **20**, 1–4.
- Carmean, D., and Crespi, B. J. (1995). Do long-branches attract flies? *Nature* **373**, 666.
- Dogiel, V. A. (1965). "General Protozoology," 2nd ed. Oxford Univ. Press, Oxford, UK.
- Edwards, A. W. F. (1992). "Likelihood," expanded ed. Johns Hopkins Univ. Press, Baltimore.
- Farris, J. S., Albert, V. A., Källersjö, M., Lipscomb, D., and Kluge, A. G. (1996). Parsimony jackknifing outperforms neighbor-joining. *Cladistics* **12**, 99–124.
- Felsenstein, J. (1973). Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Syst. Zool.* **22**, 240–249.
- Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27**, 401–410.
- Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* **17**, 368–376.
- Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**, 783–791.
- Felsenstein, J. (1988). Phylogenies from molecular sequences: Inference and reliability. *Annu. Rev. Genet.* **22**, 521–565.
- Gaut, B. S., and Lewis, P. O. (1995). Success of maximum likelihood in the four-taxon case. *Mol. Biol. Evol.* **12**, 152–162.
- Goldman, N. (1990). Maximum likelihood inference of phylogenetic trees, with special reference to a Poisson process model of DNA substitution and to parsimony analyses. *Syst. Zool.* **39**, 345–361.
- Grasse, P.-P., and Lavette, A. (1978). La myxosporidie *Sphaeromyxa sabrazei* et le nouvel embranchement de Myxosporidies (Myxozoa). Recherches sur l'état pluricellulaire primitif et considerations phylogenetiques. *Ann. Sci. Nat. Zool. (Paris)* **20**, 193–285.
- Hadzi, J. (1963). "The Evolution of the Metazoa." MacMillan Co., New York, NY.
- Hanelt, B., Van Schyndel, D., Adema, C. M., Lewis, L., and Loker, E. S. (1996). The phylogenetic position of *Rhopalura ophiocoma* (Orthonectida) based on 18S ribosomal DNA sequence analysis. *Mol. Biol. Evol.* **13**, 1187–1191.
- Hillis, D. M. (1995). Approaches for assessing phylogenetic accuracy. *Syst. Biol.* **44**, 3–16.
- Hillis, D. M., Huelsenbeck, J. P., and Swofford, D. L. (1994). Consistency: Hobgoblin of phylogenetics? *Nature* **369**, 363–364.
- Huelsenbeck, J. P. (1995). Performance of phylogenetic methods in simulation. *Syst. Biol.* **44**, 17–48.
- Huelsenbeck, J. P. (1997). Is the Felsenstein zone a fly trap? *Syst. Biol.* **46**, 69–74.
- Huelsenbeck, J. P., and Hillis, D. M. (1993). Success of phylogenetic methods in the four-taxon case. *Syst. Biol.* **42**, 247–264.
- Huelsenbeck, J. P., Hillis, D. M., and Jones, R. (1996). Parametric bootstrapping in molecular phylogenetics: Applications and performance. In "Molecular Zoology" (J. D. Ferraris and S. R. Palumbi, Eds.), pp. 19–45. Wiley-Liss, New York.
- Kathirithamby, J. (1989). Review of the order Strepsiptera. *Syst. Entomol.* **14**, 41–92.
- Kinzelbach, R. K. (1971). Morphologische Befunde an Fächerflüglern und ihre phylogenetische bedeutung (Insecta: Strepsiptera). *Zoologica* **119**(1/2), 1–256.
- Kinzelbach, R. K. (1990). The systematic position of Strepsiptera (Insecta). *Am. Entomol.* **36**, 292–303.
- Kristensen, N. P. (1981). Phylogeny of insect orders. *Annu. Rev. Entomol.* **26**, 135–157.
- Kristensen, N. P. (1991). Phylogeny of extant hexapods. In "The insects of Australia: A textbook for students and research workers" (I. D. Naumann, P. B. Carne, J. F. Lawrence, E. S. Nielsen, J. P. Spradberry, R. W. Taylor, M. J. Whitten, and M. J. Littlejohn, Eds.), 2nd ed., pp. 125–140. CSIRO, Melbourne Univ. Press, Melbourne.
- Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In "Criticism and the growth of knowledge" (I. Lakatos and A. Musgrave, Eds.), pp. 91–196. Cambridge Univ. Press, Cambridge.
- Lom, J. (1990). Phylum Myxozoa. In "Handbook of Protoctista" (L. Margulis, J. O. Corliss, M. Melkonian, and D. J. Chapman, Eds.), pp. 36–52. Jones and Bartlett Publishers, Boston, Massachusetts.
- Pawlowski, J., Montoya-Nurgos, J. -I., Fahrni, J. -F., Wüest, J., and Zaninetti, L. (1996). Origin of the Mesozoa inferred from 18S rRNA gene sequences. *Mol. Biol. Evol.* **13**, 1128–1132.
- Picken, L. E. R., and Skaer, R. J. (1966). A review of researches on nematocysts. In "The Cnidaria and Their Evolution" (W. J. Rees, Ed.), pp. 19–50. Zool. Soc. London, London, UK.
- Pix, W., Nalbach, G., and Zeil, J. (1993). Strepsipteran forewings are haltere-like organs of equilibrium. *Naturwissenschaften* **80**, 371–374.
- Popper, K. (1983). "Realism and the Aim of Science." Routledge, London.
- Shostak, S. (1993). A symbiogenic theory for the origins of cnidocysts in Cnidaria. *Biosystems* **29**, 49–58.
- Siddall, M. E. (1995). Another monophyly index: Revisiting the jackknife. *Cladistics* **11**, 33–56.
- Siddall, M. E. (1997). Prior agreement: Arbitration or arbitrary. *Syst. Biol.* **46**, 766–770.
- Siddall, M. E. (1998). Success of parsimony in the four-taxon case: Long-branch repulsion by likelihood in the Farris Zone. *Cladistics* **14**, 209–220.
- Siddall, M. E., and Bureson, E. M. (1998). Phylogeny of leeches (Hirudinea) based on mitochondrial cytochrome *c* oxidase subunit I. *Mol. Phylog. Evol.* **9**, 156–162.
- Siddall, M. E., and Kluge, A. G. (1997). Probabilism and phylogenetic inference. *Cladistics* **13**, 313–336.
- Siddall, M. E., Martin, D. S., Bridge, D., Cone, D. M., and Desser, S. S. (1995). The demise of a phylum of protists: Myxozoa and other parasitic Cnidaria. *J. Parasitol.* **81**, 961–967.
- Smothers, J. F., Von Dohlen, C. D., Smith, L. H. Jr., and Spall, R. D. (1994). Molecular evidence that the myxozoan protists are metazoans. *Science* **265**, 1719–1721.
- Stolc, A. (1899). Actinomyxidies, nouveau groupe de Mesozoaires parent des Myxosporidies. *Bull. Intl. Acad. Sci. Boheme*. **22**, 1–12.

- Swofford, D. L. (*forthcoming*) PAUP\*—Phylogenetic Analysis Using Parsimony (\* *inter alia*).
- Swofford, D. L., Olsen, G. J., Waddell, P. J., and Hillis, D. M. (1996). Phylogenetic inference. In "Molecular Systematics" (D. M. Hillis, C. Moritz, and B. K. Mable, Eds.), pp. 407–514. Sinauer Associates, Sunderland, MA.
- Weill, R. (1938). L'interpretation des Cnidosporeidies et la valeur taxonomique de leur cnidome. Leur cycle compare a la phase larvaire des Narcomeduses Cuninides. *Trav. Stat. Zool. Wimereaux* **13**, 727–744.
- Wenzel, J. J., and Siddall, M. E. (1999). Noise. *Cladistics* **15**, 51–64.
- Wheeler, W. C. (1995). Sequence alignment, parameter sensitivity, and the phylogenetic analysis of molecular data. *Syst. Biol.* **44**, 321–331.
- Whiting, M. F. (1998a). Phylogenetic position of Strepsiptera: Review of molecular and morphological evidence. *Int. J. Insect Embryol. Dev.* **27**, 53–60.
- Whiting, M. F. (1998b). Long-branch distraction and the Strepsiptera. *Syst. Biol.* **47**, 134–138.
- Whiting, M. F., and Kathirithamby, J. (1995). Strepsiptera do not share hind-wing venational synapomorphies with the Coleoptera. *J. N. Y. Entomol. Soc.* **103**, 1–14.
- Whiting, M. F., and Wheeler, W. C. (1994). Insect homeotic transformation. *Nature* **368**:696.
- Whiting, M. F., Carpenter, J. C., Wheeler, Q. D., and Wheeler, W. C. (1997). The Strepsiptera problem: Phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. *Syst. Biol.* **46**, 1–68.