



# Success of Parsimony in the Four-Taxon Case: Long-Branch Repulsion by Likelihood in the Farris Zone

Mark E. Siddall

*Laboratory of Phylohirudinology, Museum of Zoology, University of Michigan, Ann Arbor, Michigan 48109,  
U.S.A.*

Received for publication 25 February 1998

---

The accuracy of phylogenetic methods is reinvestigated for the four-taxon case with a two-edge rate and a three-edge rate. Unlike previous studies involving computer simulations, the two-edge rate relates to branches that are sister taxa in the model tree. As with previous studies, certain methods are found to behave inaccurately in a portion of the parameter space where the two-edge rate is proportionally large. This phenomenon, to which parsimony is immune, is termed “long-branch repulsion” and the region of poor performance is called the Farris Zone. Maximum likelihood methods are shown to be particularly prone to failure when closely related taxa have long branches. Long-branch repulsion is demonstrated with an empirical case involving Strepsiptera and Diptera.

© 1998 The Willi Hennig Society

---

## INTRODUCTION

Molecular systematists may choose among many methods of phylogenetic inference. These include

---

Correspondence to: Mark Siddall. E-mail: [msiddall@umich.edu](mailto:msiddall@umich.edu);  
Fax: (313) 663 4080.

parsimony (Hennig, 1950; Kluge and Farris, 1969; Farris et al., 1970), phenetics (Sokal and Michener, 1958), neighbor-joining (Saitou and Nei, 1987), genetic distances (e.g. Fitch and Margoliash, 1967), or maximum likelihood (Felsenstein, 1973, 1981a). The ability of many of these methods to achieve their stated goals has already been abundantly countered, whereas others have not. For example, phenetic algorithms like UPGMA explicitly require that rates of change do not differ markedly across lineages on a tree if degree of similarity is to be a measure of phylogenetic relatedness. And so, in the absence of clock-like behavior, the data more often than not contravene a primary premise of UPGMA. Distance methods, while not necessarily requiring equal rates of change, still require metricity. In so far as data may violate the requisite triangle inequality or result in negative branch lengths, again the fundamental premise for these methods denies their legitimacy (Farris, 1981, 1985, 1986). Some methods, like neighbor-joining, although fast, always yield singular solutions to data that may properly have multiple optima. Phylogeneticists should be concerned that their data lack the power to resolve relationships and it is unfortunate that neighbor-joining methods are so prone to this sort of over-resolution (Farris et al., 1996).

Ontology often is overlooked in contemporary systematics. That is, the justification of a method must logically follow from what it claims to achieve. Alphabetical ordering, for example, would render a very stable classification, and even a hierarchical tree structure, yet this is not used because it is not thought that the spellings of organisms are antecedent to discovering their historical relationships. By extension, for example, though phylogeny may cause related taxa to look on-the-whole similar, overall similarity is not antecedent to phylogeny itself. Neighbor-joining is best thought of as a way to get to an efficient preliminary estimate of an optimal tree, as opposed to an optimal tree *per se*. But then there are already methods of arriving at estimates of optimal trees that are more efficient (e.g. Farris, 1970). Ultimately, for a method to be minimally justified, it must claim to be able to *explain* character distributions. Presently, only two methods allow this: parsimony and maximum likelihood. Thus, only these two methods may be legitimately considered “phylogenetic” as opposed to mere mathematical clustering algorithms.

Parsimony explains character distributions across taxa in terms of what is minimally required by those observations, that is, its task is only to explain observed information (Farris, 1983; Kluge, 1997). It is possible, however, that there are unobserved changes in data, yet these are conjectural propositions and not observations requiring explanation by parsimony. Maximum likelihood, in contrast, is sensitive to the notion that all of the observed data may not comprise all of the change implied by it and, rather than seeking the best *explanation* of the data, seeks that estimate that is most *probable*. Given two taxa with the same state (e.g. adenosine), parsimony states that, without any other contravening change, grouping those two taxa together accounts for the distribution of adenosine with a single synapomorphous explanation of homology. Maximum likelihood, on the other hand, states that under certain circumstances, it is more probable that the two taxa acquired their adenosines independently than is the postulate that there was only one change (Swofford et al., 1996).

The proponents of parsimony and of likelihood each claim superiority for their method. More often than not, cladists direct attention to the philosophical underpinnings of their science (e.g. Platnick, 1979;

Farris, 1983; Sober 1983; Frost and Kluge, 1994; Kluge, 1997; Siddall and Kluge, 1997), whereas likelihoodists point to the sophistication of their algorithms and the associated power that comes with being able to accommodate a multitude of parameters and contingencies (e.g. Goldman, 1990; Penny et al., 1992; Huelsenbeck and Crandall, 1997). On the negative side, cladists outline what they see to be dangerous claims that one somehow knows how evolution works, whereas likelihoodists level the charge that cladists fail to make their assumptions explicit, and that one of those assumptions is that rates of change are low (e.g. Felsenstein, 1973, 1981b; Goldman, 1990; Swofford et al., 1996). Recently there has been a redirection which has intensified the debate. This new perspective comes in the form of computer simulations of phylogenetic trees in which one knows the true tree and can alter various rates of change on lineages, later assessing which methods perform better under those conditions. These simulations usually have had one basic form (e.g. Huelsenbeck and Hillis, 1993; Huelsenbeck, 1995). A five-edge tree (with four terminal taxa) is simulated according to various stochastic parameters, the most important of which are two rate parameters (P and Q in Felsenstein's (1978) original description of the problem, but later A and B or “three-branch length” and “two-branch length” in Huelsenbeck's (1995) fuller consideration) which relate to particular branches of the tree (Fig. 1). The two-branch rate obtains for two unrelated terminal taxa whereas the three-branch rate is applied to the other two unrelated terminals and the one internal interval. Huelsenbeck (1995) has argued assertively, and rather consistently (e.g. Huelsenbeck, 1997; Huelsenbeck and Crandall, 1997; Huelsenbeck and Rannala, 1997), that these simulations demonstrate the superiority of likelihood methods over parsimony because, when extended to simulated data sets of increasing numbers of characters, the likelihood methods perform better across the full parameter space than does parsimony. There is an attendant claim that, because parsimony performs so poorly under the conditions in which unrelated taxa have high rates of change, its use entails an assumption that these situations do not exist (e.g., Felsenstein, 1981b; Goldman, 1990). Even if a cladist asserts that the case considered is a very limiting one, s/he must also admit that in that limiting case parsimony could

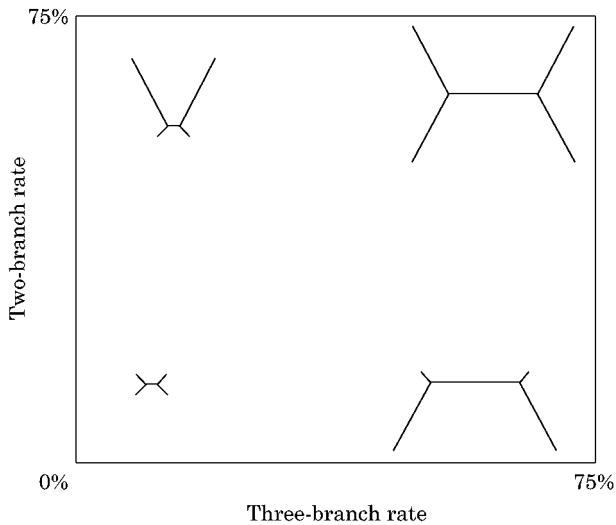


FIG. 1. Conditions previously employed in four-taxon data simulations in which the two-branch length on the ordinate relates to two unrelated terminal branches.

fail and that likelihood may not if afforded sufficient data.

To date cladists have rallied behind the assertion that one cannot know when such a limiting case exists but have shied away from having likelihoodists admit that it is nearly impossible to arrive at an efficient likelihood tree in a reasonable amount of time for the number of taxa (> four) that we regularly concern ourselves with. It is this grounding in philosophy as opposed to naked operationalism that has spared likelihoodists a most potent attack on their methods. What is difficult to explain is why more have not seen this critique of parsimony for what it most assuredly is: a limiting case. Being more concerned with explanatory power than appeals to a truth which one cannot know, and being unwilling to invoke processes that one cannot countenance, a cladist must admit that if two unrelated long-branched taxa are separated by a short internode (in truth), the choice of parsimony might be misleading, even though one cannot know this for certain or with any degree of certainty for a particular case. What has not yet been demonstrated by those who seek to justify methods according to limiting cases, is whether or not there might be some other limiting case in which parsimony is guaranteed to get the right answer, but in which likelihood is bound to fail. If such a case is discovered, would those who have tried to extract admissions from cladists not then

admit that the use of likelihood assumes that this limiting case does not obtain, even if they, too, have not been explicit about it?

Following the rubric that simulated phylogenies are a reasonable way to determine the merits of competing methods (but see [Edwards, 1995](#)) I will show here that such a limiting case does exist. The confounding phenomenon that likelihood is shown to be exceptionally sensitive to is “long-branch repulsion”. Similar to previous methods in which a three dimensional performance space has been defined, there is a region of inefficiency that delimits the conditions in which likelihood methods more often than not will fail to converge on the correct model topology: the Farris Zone.

## METHODS

The general approach taken in this study was to construct DNA sequence data for a specific model tree using computer simulation and then evaluate whether parsimony, likelihood, neighbor-joining or distance methods estimate the “correct” tree (the tree on which the data were generated). The similarity of description of methods here with those of [Huelsenbeck \(1995\)](#) should be obvious.

An unrooted four-taxon tree was used in this study (Fig. 2). This tree has one internal branch and four terminal branches. Every simulated tree was constructed from two branch-length parameters. One

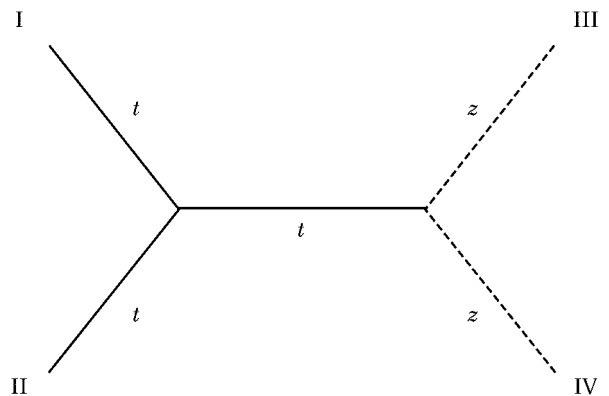


FIG. 2. The unrooted four-taxon tree used as the model phylogeny in this study. Branches labeled  $t$  were constrained to be of equal length as were the branches labeled  $z$ . The two-branch length ( $z$ ), however, related to sister taxa III and IV.

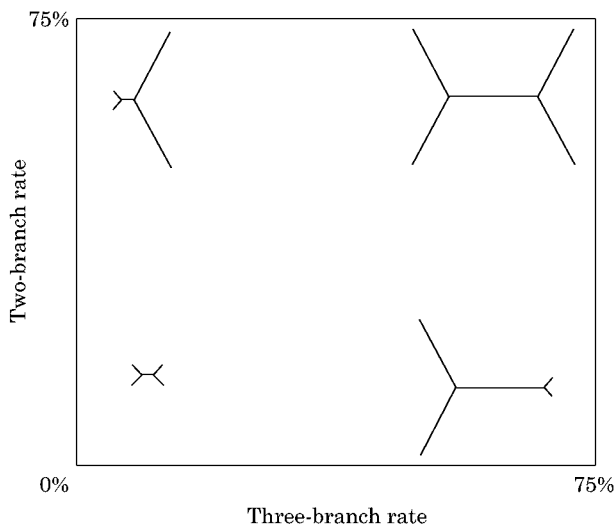


FIG. 3. Parameter space examined for the four-taxon model tree. The two-branch length is on the ordinate and the three-branch length is on the abscissa. Amounts of expected internodal change ranged from 0.75 to 75% in increments of 7.5%.

parameter established the expected percentage change of the internal branch and two adjacent terminal branches ( $t$  branches in Fig. 2, here called the three-branch length). The other parameter established the expected percentage change of the remaining two adjacent terminal branches ( $z$  branches in Fig. 2, here called the two-branch length). In light of the unavailability of processing power afforded by platforms like a Connection Machine™ (CM-5), and the need to compute likelihood trees with PAUP\* (Swofford, forthcoming) on a platform that most researchers also will have available to them, it would have been impossible to fully explore all intermediate values for which these parameters can vary. By way of compromise, instead of assessing 2% increments of expected internodal change, the constrained branches were varied in increments of approximately 7.5%, from 0.75 to 75% (Fig. 3). Each combination of  $t$  and  $z$  was simulated 100 times.

Huelsenbeck (1995: 22) remarked that if all five branches were allowed to vary, the number of required simulations would be prohibitively large and that results in five dimensions would be difficult to visualize. This placed a limitation on the kind of branch length variation that Huelsenbeck (1995) could assess, and he restricted his treatment to the specific case in which parsimony had been predicted to fail. Here, again, it would be difficult to perform simula-

tions for multiple dimensions, but future work might focus on those other dimensions. Whereas I focus here on one possible two-rate parameter space and Huelsenbeck and Hillis (1993; and Huelsenbeck, 1995) focused on another, there is one additional two-rate parameter space where two edges have the same rate, that is, a terminal branch and an internal branch have one rate and the remaining three terminal branches have another. As well, there are two additional two-rate parameter questions in which four branches have one rate and one branch has another. Yang (1996) pursued some of these alternatives and found parsimony to be more efficient than likelihood methods.

One potential criticism of the four-taxon parameter space is whether it should be considered reasonable in terms of some of the implied amounts of change. However, these ranges are the same as those employed by Huelsenbeck (1995: 22–23) in his critique of parsimony methods and he provided ample justification for these rates of change therein (as did Hillis and Huelsenbeck, 1995; but see Nei et al., 1995).

The simplest model of DNA substitution was assumed in this study. The Jukes–Cantor (1969) model has a parameter  $\lambda$  that determines the rate at which a given nucleotide changes to one of the three remaining nucleotides. More complex mathematical formulae illustrating the model can be found elsewhere (e.g. Jukes and Cantor, 1969; Huelsenbeck, 1995); however, it merely amounts to a statement that if a nucleotide changes, it has equal probability of changing to one of the other three nucleotides. Whether or not the site will change on a particular branch is determined by the rate parameters  $t$  and  $z$ .

Once constructed, the simulated data were subjected to analysis by parsimony, neighbor-joining (nj), least-squares distance analysis with an appropriate to the model Jukes–Cantor correction [d(JC)], (I have not used the term “minimum evolution” for what has traditionally been called “distance” because all phylogenetic methods are a form of minimization in some sense or another and I remain convinced that obfuscation is unhelpful) and by several constraining models of maximum likelihood (ml). In parsimony inferences, all sites were unweighted. The four models assumed in likelihood estimations included a simple Jukes–Cantor (1969) model [ml(JC)]; a Kimura

(1980) two-parameter model [ml(K2P)] in which transition:transversion ratios were estimated from the simulated data; the Felsenstein (1981a) model [ml(F81)] in which nucleotide compositions were estimated from the data; and the Hasegawa–Kishino–Yano (1985) model [ml(HKY)] in which both transition:transversion ratios and base composition parameters were estimated from the data. Ostensibly the more complex models, ml(K2P), ml(F81), and ml(HKY), should each approximate a Jukes–Cantor model of change because the additional parameters are estimated from data modeled according to this process. However, stochastic variation away from a strict Jukes–Cantor process can be expected and the methods that have additional parameters might be expected to perform better than the simplest one.

The performance of phylogenetic methods was determined by finding how frequently the modeled tree was recovered in resulting analyses. In order to dispose these values favorably towards indecision of the method, only those cases in which the correct solution was *not* found counted against these frequencies. Thus, if multiple trees were found to be optimal, and at least one of those trees was the correct tree, the method was considered successful in finding the correct tree, notwithstanding that the method was indecisive.

## RESULTS

The frequency with which each method uncovered the correct model tree for 100, 500 and 1000 characters is illustrated in Fig. 4, with the three-branch length on the abscissa and the two-branch length on the ordinate. The diagonal represents equal branch lengths. Color varies from dark blue (100% recovery of the “correct” tree) to red (30% recovery of the “correct” tree). In addition, the 50% and 95% isoclines for recovery are illustrated for the case in which 1000 sites were allowed to vary.

For all simulations, parsimony performed much better than the other methods examined. In simulations which used 100 characters there was an indication of less than perfect performance in the lower left region of the parameter space, where rates of change are equal but very low. In this case, recovery of an incorrect topology occurred with a frequency of about 20%. Similarly, in the upper right hand region of the

parameter space, with only 100 characters, where rates of change are equal but high, the correct topology was recovered in slightly more than half of the simulations. With an increase to 500 or 1000 variable characters, there was enough change in the equal-but-low rate simulations to always correctly recover the modeled tree. With this increased number of characters, in the case of equal but high rates, recovery frequencies exceeded 95%.

Except for parsimony, all of the methods performed poorly in the upper left hand corner of each plot (Fig. 4), a region in which two long-branched taxa are each others' closest relatives and the remaining three branches in the modeled tree are comparatively short. This region is termed the Farris Zone because Farris (1983: 682) noted that any method that performs well under one set of circumstances is bound to perform poorly under others, one “need only imagine that his characters have evolved in just the right way to lead him to a false conclusion”. Where rates of change were proportionally high for the two-branch length, parsimony always recovered the model tree.

Likelihood methods all performed rather poorly in the Farris Zone, though they otherwise recovered the correct tree most of the time across much of the parameter space. As the two-branch rate (for sister taxa) was made progressively larger, the ability of likelihood methods to recover the correct topology became worse. With fewer variable characters (i.e. 100) the use of a JC or F81 model in likelihood resulted in accuracies that were lower than use of a K2P model or the HKY model, notwithstanding that the JC model was appropriate to the modeled process of nucleotide change. With only 100 available characters, employing a JC or F81 model of change approached a level of accuracy of about 45% in the Farris Zone whereas the K2P and HKY models each approached accuracies of about 55%. However, as the number of characters was increased to 500 or 1000, the relative accuracies of all implementations of likelihood varied around 33% which is equivalent to randomly picking one of the three possible topologies for four taxa. The likelihood methods also recovered the correct topology with somewhat lower accuracies than parsimony when all branch rates were equal but high (in the upper right hand region of the plots). This same phenomenon was evident in Huelsenbeck's

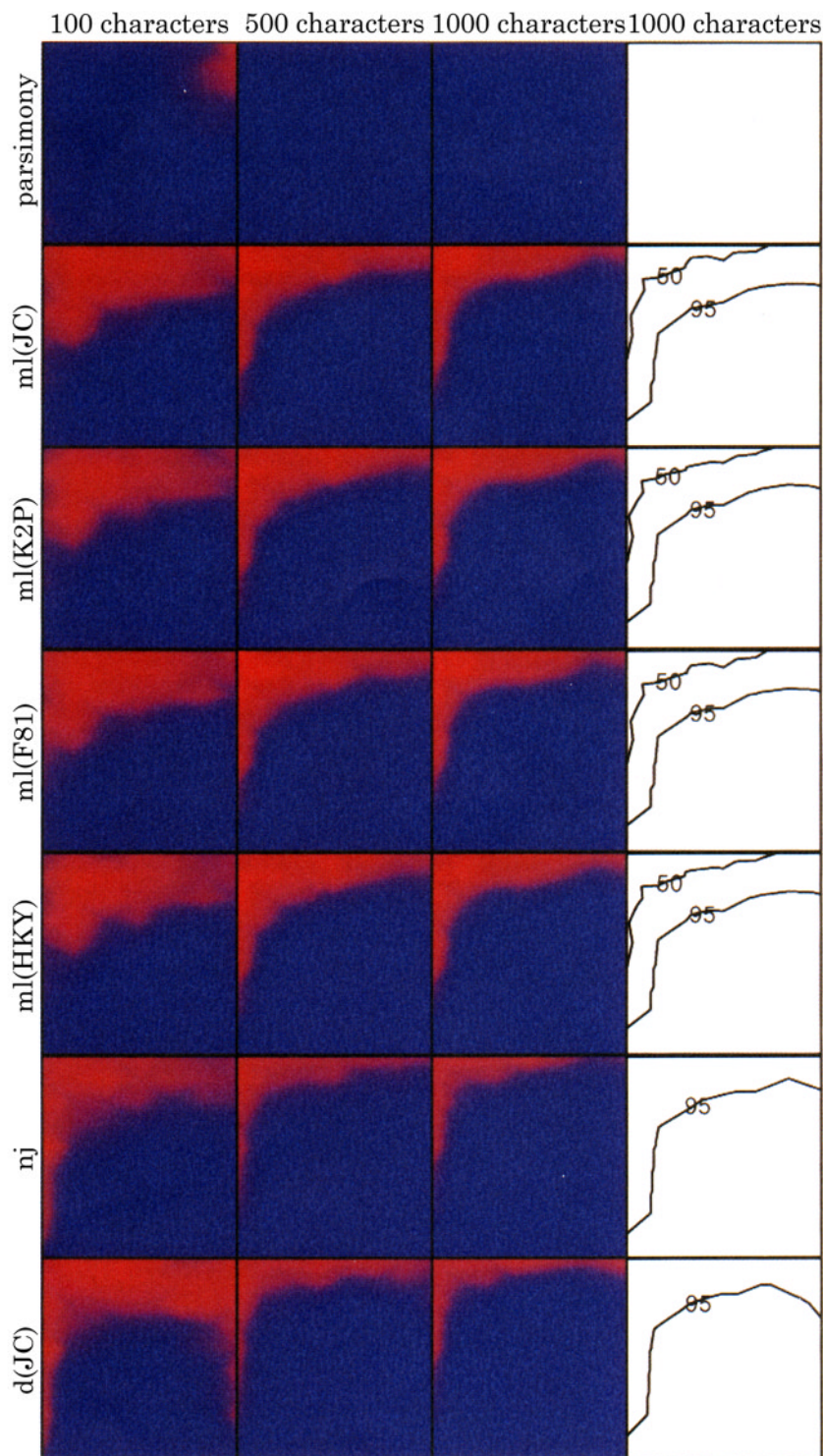


FIG. 4. The frequency of estimating the correct model tree for seven phylogenetic methods in the parameter space outlined in Fig. 3. Spline interpolated values of performance range from dark blue 100% recovery, to red 30% recovery, when 100, 500 or 1000 characters are free to vary. In addition, the 50% and 95% frequency contour lines are plotted for the case involving 1000 characters. These contours are absent for parsimony because, with 1000 characters free to vary, this method reconstructed the correct model tree more than 95% of the time across the whole parameter space. Similarly, the 50% contour lines are absent for nj and d(JC) plots. ml, maximum likelihood; nj, neighbor-joining; d, distance a.k.a. "minimum evolution"; JC, Jukes] Cantor 1969. model; K2P, Kimura 1980. two-parameter model; F81, Felsenstein 1981a. model; HKY, Hasegawa] Kishino] Yano 1985. model.

(1995) simulations but was not noted at that time.

The neighbor-joining and distance methods each recovered the correct tree with about equal frequency across the parameter space for varying numbers of informative characters, and with slightly better performance than the likelihood methods when the two-branch rate ( $z$ ) was made large. That is, although likelihood methods picked a tree at random in the Farris Zone with 500 characters, nj and d(JC) recovered the correct tree with slightly better frequencies (approx. 50%).

## DISCUSSION

Much has been made of the poor performance of parsimony in what has been termed “the” four-taxon case, yet, as is evident from the simulations investigated here, that case is not a particularly general one and is not the only four-taxon case conceivable. [Felsenstein \(1978\)](#) postulated that when long-branched taxa are separated by a short internal branch, parsimony would group the long branches together. This phenomenon has been referred to as long-branch attraction and has been characterized as a “hobgoblin” for parsimony ([Hillis et al., 1994](#)). Likelihood methods have been shown to recover the true modeled tree more readily in cases when long-branched lineages are not sister taxa. The most troubling result is that in the corollary situation where long-branched taxa are sisters, likelihood methods fail so regularly while parsimony routinely recovers the correct tree. [Felsenstein \(1978: 408, and 1973, citing Wald, 1949\)](#) has claimed that likelihood methods “have desirable statistical properties such as consistency and asymptotic efficiency”. But this cannot be true. Among [Wald’s \(1949\)](#) criteria for consistency were requirements for independence and identical distributions, which sequenced nucleotides cannot have, and that the likelihood function is everywhere continuous and continuously differentiable with respect to the parameter of interest. Cladograms being discrete, it has yet to be explained how that condition can be satisfied or indeed what it would mean in this case ([Farris, 1997](#)). With more than 100 characters, likelihood methods did actually choose one of the topologies in all but a very small fraction

of the simulations, and did not equivocate among multiple equally optimal solutions. That is, if likelihood scores were equal among the three topologies, these results would not be so damning of those methods, notwithstanding that relative performance frequencies would still be about 33%. However, all cases in which the correct topology was recovered, even if there were multiple topologies, were considered as counting in favor of the method here. The Farris Zone of poor recovery depicted in [Fig. 4](#), then, indicates failure to recover the correct tree *at all*. Paraphrasing [Goldman \(1990: 357\)](#), “this must cast doubt on all likelihood inferences”. On the other hand, in the case of infinite data where sister taxa are long-branched, parsimony will be “positively leading”.

It is significant that likelihood performs so poorly and parsimony so well in precisely the inverse of the situation in which parsimony performs poorly and likelihood performs well. That is, if in truth there are two long-branched taxa which are not sister taxa and they are separated by a short internal branch, parsimony will tend to group them together artificially (long-branch attraction) whereas likelihood will tend to keep them apart. But if the two long-branched taxa are, after all, sister taxa, then likelihood will tend to force the branches apart (long-branch repulsion) and parsimony will correctly group them together. This will present a serious quandary in any empirical case using real data from real taxa.

The great difficulty, then, is how to conduct oneself admitting that the truth is not knowable. The manifestation in any single study where parsimony groups two long branches together and likelihood does not, now can only be taken to be indicative of a discrepancy in methods. Such a finding (e.g. [Huelsenbeck, 1997](#)) leaves the question open as to whether parsimony is causing the taxa to group together or likelihood is forcing them apart. Thus, long-branch attraction cannot be proven empirically nor invoked as a rationale for using one method or another. Nor, for that matter, can long-branch repulsion.

It is instructive, in such a quandary, to consider what evidence there is, and what is not evidence. For the parsimony solution to be adopted, one need only take the evidence of character generalities as they are. No stochastic model need be assumed, and it is clear that the charge that parsimony assumes low rates of

change (Felsenstein, 1973, 1981b, 1983; Felsenstein and Sober, 1986; Goldman, 1990) is false. That is, in the upper left region of the plots (Fig. 4), rates of change are high for two sister taxa and very low for others but parsimony recovers the correct tree with 100% frequencies. The reason that this is so is that the number of synapomorphies recovered for a pair of sister taxa need not all actually be homologies for the method to have behaved correctly (Farris, 1983: 680). In contrast, for the likelihood solution to be adopted, a stochastic model must be invoked and yet justification for the model cannot be found within the method itself (Edwards, 1972; Goldman, 1990). There must also be the acceptance of the notion that absence of evidence nonetheless is evidence in and of itself. Admittedly, it is quite reasonable to suppose that there have been more nucleotide changes than are minimally required, but how many and when and where on the tree they might have occurred is more problematic. This comes full circle to having to invoke some stochastic law of change and to denying the relevance of such things as selection or functional constraints. Although it is conceptually reasonable as a generalization, there never are observations in any particular case of more substitutions than are minimally required. The only direct evidence of substitutions are for those that are minimally required. Both parsimony and likelihood admit the same minimum possible number of changes (no likelihood method will arrive at a shorter tree than parsimony); additional changes are conjectures made by invoking the model in the likelihood method itself. They are not observations that stand apart from the choice of method.

There is ample reason to believe that taxa with similar rates of change are, in fact, more likely than not to be related. Branchiobdellidan annelids exhibit only half a dozen changes in their 18S rDNA for the whole order, notwithstanding that their sister group, the leeches, exhibit much faster rates in this gene than do other clitellates (Fig. 5A). Coding and non-coding sequences render longer branch lengths for the protistan phylum Haplosporidia than for related taxa (Fig. 5B). That one should really expect two or more long-branched taxa to be related has precedent in analyses of real taxa. For example, analyses of holometabolous insects by Carmean and Crespi (1995), Whiting et al. (1997), Huelsenbeck (1997) and

Whiting (1998) all show high rates of change for dipterans relative to other groups; cnidarians consistently reveal markedly reduced rates of change relative to Bilateria (Bridge et al., 1995; Siddall et al., 1995); teleost hemoglobins evolve faster than nototheniid hemoglobins (Stam et al., 1997); branch lengths for argasid ticks are longer than for ixodid ticks (Black et al., 1997); mitochondrial 12S rDNA changes faster in South American turtles than it does in Australasian turtles (Seddon et al., 1997); one paralogue in a gene family, on the whole, may demonstrate a rate of change that outpaces that of another paralogue (Swaffield and Purugganan, 1997); the taxa in well-supported groups like tracheophytes (Lewis et al., 1997) and Pyrenomycetes (Lo Buglio et al., 1996) have longer branches than their sister groups; and *Lactobacillus* species all have longer branches than *Enterococcus* species (Evers et al., 1996). Perhaps the most familiar example involves a clade that few will doubt: eukaryotic mitochondrial rRNA genes all have longer branches than the equivalent genes in bacteria from whence the former are derived (Denovan-Wright et al., 1996). So, it would seem that when faced with a result in which parsimony groups two long-branched taxa together, and likelihood does not, we already have sufficient reason to believe the most parsimonious solution because related long branches are to be expected anyway.

Is there evidence of likelihood falling into the Farris Zone in practice? Consider the case concerning Strepsiptera and Diptera (Carmean and Crespi, 1995; Whiting et al., 1997; Huelsenbeck, 1997; Whiting, 1998). Huelsenbeck (1997) regarded the results of his analyses to be indicative of empirical proof of long-branch attraction because parsimony grouped Strepsiptera with Diptera whereas likelihood analyses (irrespective of model) did not (Fig. 6A, B). The branches leading to Strepsiptera and the two dipterans are markedly longer than are those for the other holometabolous insects in Carmean and Crespi's (1995) analysis. Huelsenbeck (1997: 69–70) argued that long-branch attraction can be invoked on the condition (among others) "that a method that is less sensitive to the long-branch attraction problem gives a phylogenetic estimate in which the long branches are separated". However, in light of the observation that this very method is prone to repulsion when sister taxa are long-branched, the issue of strep-

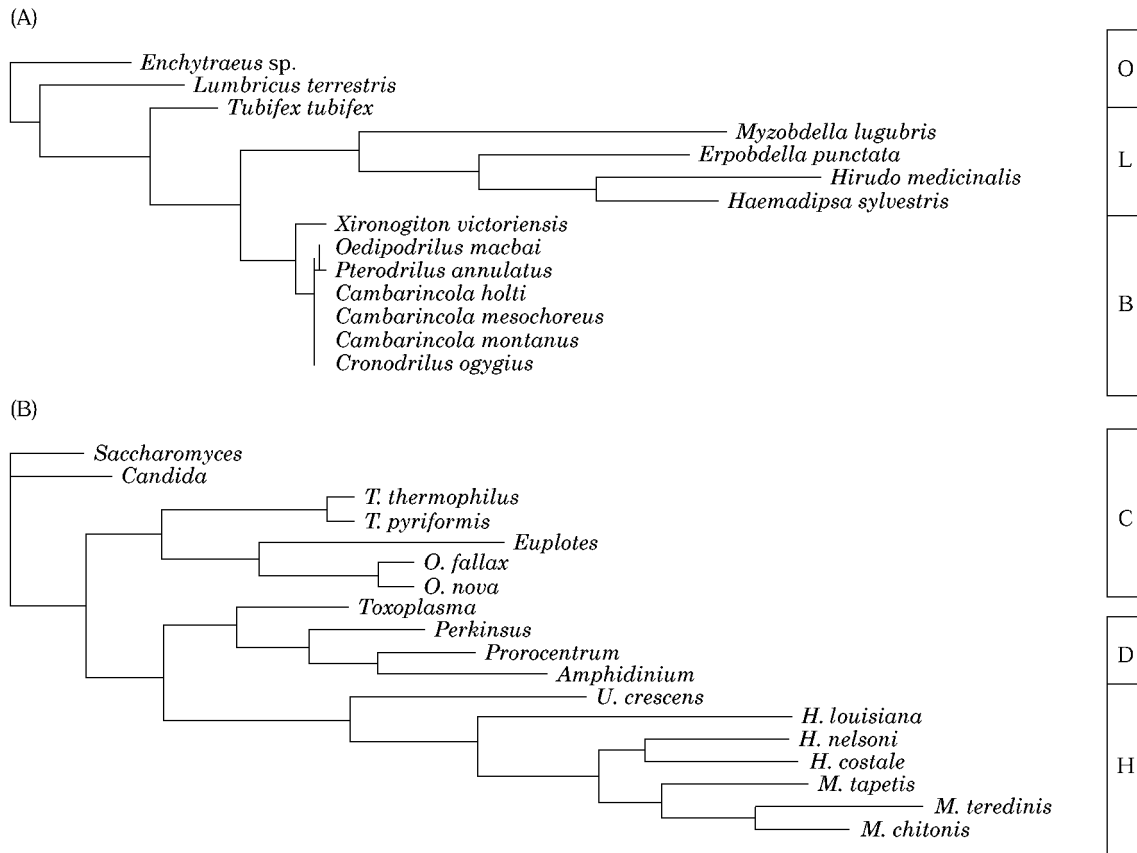


FIG. 5. Phylogenetic analyses demonstrating heritability of constraints resulting in similar rates of change for closely related taxa. (A) 18S rDNA of clitellates shows high rates of change for leeches (L) and markedly low rates of change for their sister-group, branchiobdellidans (B), relative to oligochaetes (O). (B) Combined 18S rDNA and actin sequences for alveolates indicating high rates of change for the haplosporidians (H) relative to related dinoflagellates (D) and ciliates (C).

sipteran relationships and long branches deserves further consideration. Of course, the truth of the matter is not knowable, but the relative delicacy of competing methods can be instructive. As it happens, likelihood methods only separate Strepsiptera from Diptera when two-thirds of the data are omitted (i.e. those excluded by Huelsenbeck, 1997). Inclusion of all of the 18S rDNA characters in that data set groups them together even in likelihood analyses (Fig. 6C). Moreover, if but one additional character (#203 in the alignment) is omitted along with the 1471 omitted in Huelsenbeck's (1997) analyses, likelihood returns a Strepsiptera + Diptera clade (Fig. 6D) much as parsimony continues to do. Whiting et al. (1997) demonstrated that more comprehensive analyses including 52 taxa (instead of the 15 used by Huelsenbeck, 1997) and three data sets [1,116 18S rDNA characters, 398 28S rDNA characters and 176 morpho-

logical characters, as opposed to the 834 sites used by Huelsenbeck (1997)], independently and in combination confirm a sister-group relationship for Strepsiptera and Diptera. Siddall and Whiting (1997) have shown that the likelihood analyses which separate strepsipterans from dipterans are prone to taxonomic instability and to alignment instability, whereas parsimony analyses unequivocally and consistently place these two groups together irrespective of taxonomic inclusion and with those same slight differences in alignment. It would appear that rather than being indicative of a proof of long-branch length attraction by parsimony, in retrospect, Huelsenbeck's (1997) analyses are demonstrative of long-branch repulsion by likelihood, though the phenomenon was not recognized at the time.

Simulation studies have been taken to represent a "first cut" for the potential usefulness of phyloge-

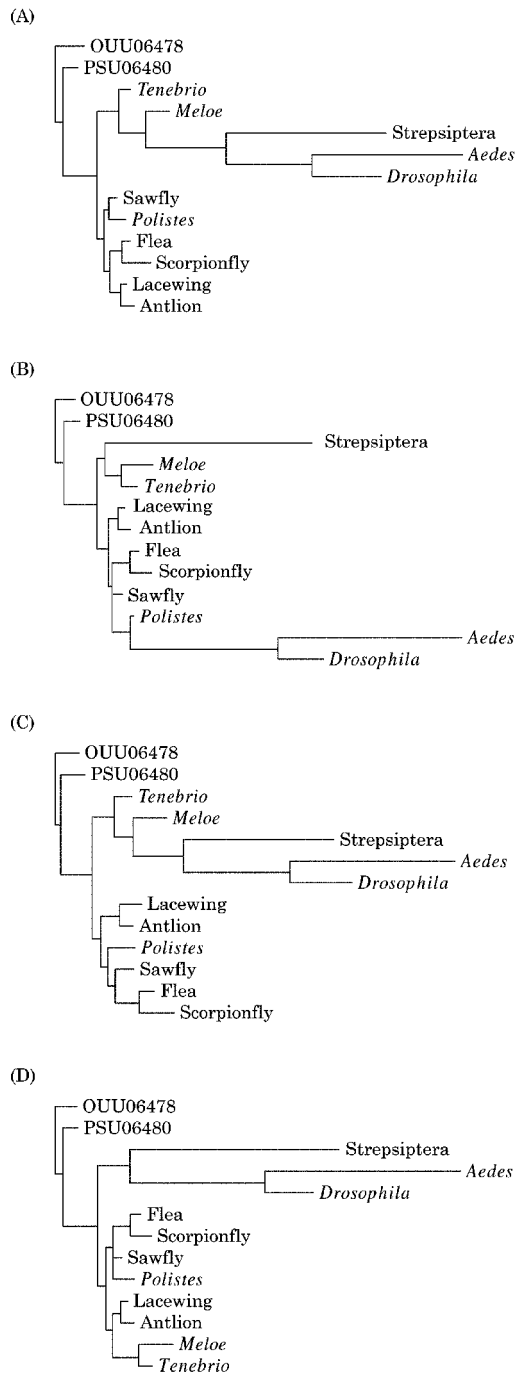


FIG. 6. Likelihood methods exhibit long-branch repulsion in analyses of Carmean and Crespi's (1995) holometabolan 18S rDNA data. With two-thirds of the data excluded (*sensu* Huelsenbeck, 1997), parsimony analysis (A) groups Strepsiptera with the two dipterans but likelihood (B) forces these taxa apart (see also Huelsenbeck, 1997). When all of the data are included (C) or if one additional character is excluded (D), likelihood methods also group Strepsiptera with Diptera.

netic methods (Huelsenbeck, 1995). It has been claimed that if a method does not perform well under very simple models of evolutionary change or if a method does not perform well even when all of its assumptions are satisfied, that method might be thought to have little hope of performing well for real character data (Huelsenbeck, 1995). Edwards (1995:253), however, disputed such claims, arguing that simulations are "valuable in assessing the robustness of a model, but not the suitability of an estimation procedure". Whatever the case, it is clear from the simulations performed here that there is no reason to suppose that parsimony requires slow rates of change any more than does likelihood. If the charge is to be leveled at parsimony that it presupposes low rates of change in unrelated taxa then an equally damaging charge must be leveled at likelihood: that it presupposes low rates of change in related taxa. Very little of substance is likely to flow from such a debate.

Rather than taking the results of simulation studies such as those performed here and elsewhere (e.g. Huelsenbeck and Hillis, 1993; Huelsenbeck, 1995) as a panacea for the difficulties posed by the choice among competing methods, phylogeneticists should be more concerned with the basic foundations of their methods. There are two philosophies of science at play here and more generally in empirical fields: verification and refutation. Phylogenetics is no more immune to their precepts than are other disciplines. How these philosophies might impinge on our choice of methods has been detailed elsewhere (e.g. Siddall and Kluge, 1997). How much one is willing to assume, and the degree with which one is willing to manipulate basic observations and discard others like morphology and behavior, will continue to be the principle difference in choice among methods. Parsimony remains the method that retains greatest fidelity to observation, imposing the fewest assumptions about process and admitting the widest variety of information arriving in molecular, morphological, fossilized, or behavioral forms. That parsimony lacks the impressive technicalities of likelihood is a triumph of sensibility.

## NOTE ADDED IN PROOF

In a talk at the 1995 Society of Systematic Biology

meeting, David Swofford apparently reported the superior accuracy of parsimony methods over likelihood methods when sister taxa have long-branches. It is my understanding that this did not involve an examination of the full four-taxon parameter space described here. Subsequent simulations show that provided one has sequence data from whole mitochondrial genomes, all evolving according to the same stochastic process, one can expect likelihood to recover the correct topology about 70% of the time in the Farris Zone. I thank David Swofford for drawing my attention to this unlikely situation. Peter Waddell (pers. comm.) previously has referred to this region as the anti-Felsenstein Zone which is, in a sense, semantically equivalent.

## ACKNOWLEDGEMENTS

I thank Jim Carpenter, Steve Farris, Arnold Kluge, Michael Sorenson, John Wenzel, Ward Wheeler, and Michael Whiting for their encouragement and comments on this project as it developed. The notion that likelihood might exhibit long-branch repulsion came in the form of a question from Jon Coddington during a symposium at the 16th meeting of the Willi Hennig Society. Steve Farris provided a thoughtful review. The expertise of David Bay was instrumental in the production of color plots. The diligence of Kathleen Apakupakul and Jessica Light in matters empirical allowed me to conduct all of the simulations in a reasonable length of time and with greater ease. Data relating to the 18S rDNA genes of leeches and branchiobdellidans were obtained with the aid of a grant from the National Science Foundation (BIO/DEB-9615211) in collaboration with Stuart Gelder and Eugene Bureson. Those for the alveolates also were obtained with the aid of a grant from the National Science Foundation (BIO/DEB-9629487) in collaboration with Kim Reece and Eugene Bureson. Those data may be obtained via <http://limnatis.ummz.lsa.umich.edu/pub/farris/>.

## REFERENCES

- Black, W. C., Klompton, J. S., and Keirans, J. E. (1997). Phylogenetic relationships among tick subfamilies (Ixodida: Ixodidae: Argasidae) based on the 18S nuclear rDNA gene. *Mol. Phylog. Evol.* **7**, 129–144.
- 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.
- Carmean, D., and Crespi, B. J. (1995). Do long branches attract flies? *Nature* **373**, 666.
- Denovan-Wright, E. M., Sankoff, D., Spencer, D. F., and Lee, R. W. (1996). Evolution of fragmented mitochondrial ribosomal RNA genes in *Chlamydomonas*. *J. Mol. Evol.* **42**, 382–391.
- Edwards, A. W. F. (1972). "Likelihood". Cambridge University Press, Cambridge.
- Edwards, A. W. F. (1995). Assessing molecular phylogenies. *Science* **267**, 253.
- Evers, S., Casadewall, B., Charles, M., Dutka-Malen, S., Galimand, M., and Courvalin, P. (1996). Evolution of structure and substrate specificity in D-alanine: D-alanine ligases and related enzymes. *J. Mol. Evol.* **4**, 706–712.
- Farris, J. S. (1970). Methods for computing Wagner Trees. *Syst. Zool.* **19**, 83–92.
- Farris, J. S. (1981). Distance data in phylogenetic analysis. In "Advances in Cladistics" (V. A. Funk, and D. R. Brooks, Eds), Vol. 1, pp. 3–23. New York Botanical Garden, New York.
- Farris, J. S. (1983). The logical basis of phylogenetic analysis. In "Advances in Cladistics" (N. I. Platnick, and V. A. Funk, Eds), Vol. 2, pp. 7–36. Columbia University Press, New York.
- Farris, J. S. (1985). Distance data revisited. *Cladistics* **1**, 67–85.
- Farris, J. S. (1986). Distances and statistics. *Cladistics* **2**, 144–157.
- Farris, J. S. (1997). "Who, really, is a statistician?" Paper presented at the Sixteenth Meeting of The Willi Hennig Society, October 23–26, 1997. George Washington University, Washington, D.C.
- Farris, J. S., Kluge, A. G., and Eckardt, M. J. (1970). A numerical approach to phylogenetic systematics. *Syst. Zool.* **19**, 172–179.
- 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. (1981a). Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* **17**, 368–376.
- Felsenstein, J. (1981b). A likelihood approach to character weighting and what it tells us about parsimony and compatibility. *Biol. J. Linn. Soc.* **16**, 183–196.
- Felsenstein, J. (1983). Statistical inference of phylogenies. *J. R. Stat. Soc. A* **146**, 246–272.
- Felsenstein, J., and Sober, E. (1986). Parsimony and likelihood: An exchange. *Syst. Zool.* **35**, 617–626.
- Fitch, W. M., and Margoliash, E. (1967). Construction of phylogenetic trees. *Science* **155**, 279–284.
- Frost, D. R., and Kluge, A. G. (1994). A consideration of epistemology in systematic biology, with special reference to species. *Cladistics* **10**, 259–294.
- 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.
- Hasegawa, M., Kishino, H., and Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **36**, 182–198.
- Hennig, W. (1950). "Grundzüge einer Theorie der phylogenetis-

- chen Systematik". Deutscher Zentralverlag, Berlin.
- Hillis, D. M., and Huelsenbeck, J. P. (1995). Assessing molecular phylogenies. *Science* **267**, 255–256.
- Hillis, D. M., Huelsenbeck, J. P., and Swofford, D. L. (1994). 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 Crandall, K. A. (1997). Phylogeny estimation and hypothesis testing using maximum likelihood. *Rev. Ecol. Syst.* **28**, 437–466.
- 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., and Rannala, B. (1997). Phylogenetic methods come of age: Testing hypotheses in a phylogenetic context. *Science* **276**, 227–232.
- Jukes, T. H., and Cantor, C. R. (1969). Evolution of protein molecules. In "Mammalian Protein Metabolism" (H. Munro, Ed.), pp. 21–132. Academic Press, New York.
- Kimura, M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **11**, 459–468.
- Kluge, A. G. (1997). Testability and the refutation and corroboration of cladistic hypotheses. *Cladistics* **13**, 81–96.
- Kluge, A. G., and Farris, J. S. (1969). Quantitative phyletics and the evolution of anurans. *Syst. Zool.* **18**, 1–32.
- Lewis, L. A., Mishler, B. D., and Vilgalys, R. (1997). Phylogenetic relationships of the liverworts (Hepaticae), a basal embryophyte lineage, inferred from nucleotide sequence data of the chloroplast *rbcL*. *Mol. Phylog. Evol.* **7**, 377–393.
- Lo Buglio, K. F., Berbee, M. L., and Taylor, J. W. (1996). Phylogenetic origins of the asexual mycorrhizal symbiont *Cenococcum geophilum* Fr. and other mycorrhizal fungi among the Ascomycetes. *Mol. Phylog. Evol.* **6**, 287–294.
- Nei, M., Takezaki, N., and Sitnikova, T. (1995). Assessing molecular phylogenies. *Science* **267**, 253–255.
- Penny, D., Hendy, M. D., and Steel, M. A. (1992). Progress with methods for constructing evolutionary trees. *Trends Ecol. Evol.* **7**, 73–78.
- Platnick, N. I. (1979). Philosophy and the transformation of cladistics. *Syst. Zool.* **28**, 537–546.
- Saitou, N., and Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **44**, 406–425.
- Seddou, J. M., Georges, A., Baverstock, P. R., and McCord, W. (1997). Phylogenetic relationships of chelid turtles (Pleurodira: Chelidae) based on mitochondrial 12S rRNA gene sequence variation. *Mol. Phylog. Evol.* **7**, 55–61.
- Siddall, M. E., and Kluge, A. G. (1997). Probabilism and phylogenetic inference. *Cladistics* **13**, 313–336.
- Siddall, M. E., and Whiting, M. F. (1997). "Building a Better Fly Trap". Paper presented at the Sixteenth Meeting of the Willi Hennig Society, October 23–26, 1997. George Washington University, Washington, D.C.
- 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.
- Sober, E. (1983). Parsimony methods in systematics. In "Advances in Cladistics" (N. I. Platnick, and V. A. Funk, Eds), Vol. 2, pp. 37–47. Columbia Univ. Press, New York.
- Sokal, R. R., and Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *Univ. Kans. Sci. Bull.* **28**, 1409–1438.
- Stam, W. T., Beintema, J. J., D'Avino, R., Tamburrini, M., and di Prisco, G. (1997). Molecular evolution of hemoglobins of Antarctic fishes (Notothenoidei). *J. Mol. Evol.* **45**, 437–445.
- Swaffield, J. C., and Purugganan, M. D. (1997). The evolution of the conserved ATPase domain (CAD): Reconstructing the history of an ancient protein module. *J. Mol. Evol.* **45**, 549–563.
- Swofford, D. L. (forthcoming). "PAUP\*", Phylogenetic Analysis using Parsimony *Inter Alia*". Program and documentation.
- 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, Massachusetts.
- Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Stat.* **20**, 595–601.
- Whiting, M. F. (1998). Long-branch distraction and the Strepsiptera. *Syst. Biol.* **47**, 134–138.
- 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.
- Yang, Z. (1996). Phylogenetic analysis using parsimony and likelihood methods. *J. Mol. Evol.* **42**, 294–307.