Cladistics 17, 266–281 (2001)
doi:10.1006/clad.2001.0172, available online at http://www.idealibrary.com on IDEAL®

Biases in Maximum Likelihood and Parsimony: A Simulation Approach to a 10-Taxon Case

Diego Pol* and Mark E. Siddall†

**Division of Paleontology and †Division of Invertebrates, American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024*

Accepted May 20, 2001

Biases present in maximum likelihood and parsimony are investigated through a simulation study in a 10-taxon case in which several long branches coexist with short branches in the modeled topology. The performance of these methods is explored while increasing the length of the long branches with different amounts of data. Also, simulations with different taxonomic sampling schemes are examined through this study. The presence of a strong bias in parsimony is corroborated: the well-known long-branch attraction. Likelihood performance is found to be sensitive to the mere presence extreme of branch length disparity, retrieving topologies compatible with long-branch attraction and long-branch repulsion, irrespective of the correctness of the model used.

INTRODUCTION

Studies on the performance of simulated datasets have been conducted in order to analyze the performance of methods for phylogeny reconstruction under controlled conditions. These types of analyses consist of a two-step procedure. The first step is the creation of datasets according to a Markovian model of character change and a defined element of the tree space.[1] The second step is the analysis of simulated datasets using different methodologies for the posterior comparison of the retrieved trees with the true topology.

Most of the previous approaches to simulated datasets were done with the aim of producing a controlled simulation of an evolutionary process, and the analysis of the datasets under different methodologies was done in order to infer the suitability of these methods for recovering phylogenetic information (e.g., Huelsenbeck and Hillis, 1993; Huelsenbeck, 1995, p. 34). However, this approach to simulation studies is not entirely satisfactory. First, as was pointed out earlier, it is certainly feasible but not really meaningful for choosing a method to find a particular scenario in which a certain method retrieves a wrong answer (Farris, 1983; Edwards, 1995; Siddall, 1998). Any method that performs well in one set of circumstances is bound to perform poorly in others, and findings may be interpreted just as a limiting case tied to its domain of application (Farris, 1983; Kim, 1998; Siddall, 1998). Second, the choice between likelihood and parsimony relies mainly on the acceptance or rejection of the assumption that a particular stochastic model is driving the nucleotide

---

[1] The various combinations of branch-length conditions possible for a given tree (Huelsenbeck and Hillis, 1993).

0748-3007/01 $35.00

transformations along the entire sequence analyzed. Thus, simulated datasets constructed under such a model can hardly provide evidence for taking either a cladistic or a likelihood approach (Edwards, 1995).

The present analysis does not intend to simulate datasets as a proxy of an evolutionary process and measure the probability that each method has for reconstructing the pattern of such an "evolutionary" process. The aim of the present approach is, rather, the creation of stochastically modeled datasets for which we know the degree of similarity with the assumed model employed in likelihood analysis. Thus, through simulation studies it is possible to test the robustness of likelihood methods (i.e., how they perform under violations of the model assumptions) and to explore the biases (i.e., systematic errors) inherent to likelihood approaches to phylogenetic inference (Chang, 1996; Yang, 1996; Huelsenbeck, 1998; Bruno and Halpern, 1999). In parsimony analyses, since it is claimed that there is no stochastic model of character change underlying the analysis, it is not possible to explore the statistical robustness of the method. However, it is certainly possible and useful to analyze the existence of biases in the method under certain conditions (i.e., different kinds of information determined by some dataset parameters, such as base composition or other sequence disparity). Previous approaches were concerned mainly with counting the proportion of correct answers that each method retrieves, instead of the identification of the underlying biases. Low percentages of recovery of the true topology may be indicative of the extent of bias, but they say nothing about what biases are causing the misleading results. These systematic errors are problems that affect real (finite) dataset analyses and are present in methods even if statistical consistency is warranted with infinite amounts of data (Lehmann, 1983; Kuhner and Felsenstein, 1994). Thus, we will focus our attention here on the specifics of incorrect topologies, not only that they are wrong, but how they are wrong. The aim of this approach is the discovery of an association between certain values of dataset parameters and the choice of a particular type of wrong topology (i.e., biases of the methods under analysis).

Another difference from previous approaches is the dimension of the study case analyzed. The majority of past simulation analyses were restricted to a particular segment of a four-taxon tree space. One extensively analyzed case consists of a four-taxon tree (4T case), in which two terminal long branches are the sister taxa to two short terminal branches and are connected by a short internodal branch (Fig. 1A). Felsenstein (1978) was the first to study this case, noting that if characters are set according to a stochastic model of change with a fixed number of possible states (e.g., nucleotide sequences), as the number of characters and the length of the long branches increase, parsimony is statistically inconsistent, recovering the long branches as sister taxa (Fig. 1A). In contrast, a simple model of stochastic evolution in a maximum-likelihood analysis can recover the true topology. Several studies followed Felsenstein (1978), focusing on the performance of different methods in this particular 4T case and finding similar misleading results for parsimony and accurate results for maximum likelihood (e.g., Hillis and

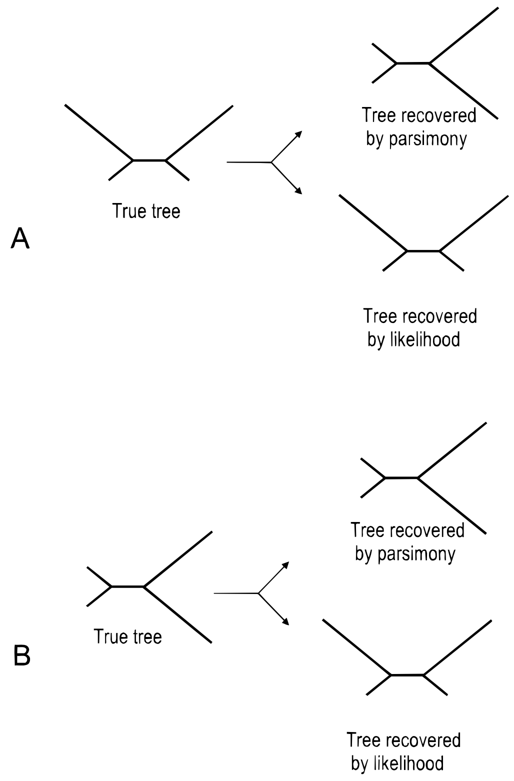

FIG. 1. Previous approaches in a 4T case. (A) The Felsenstein zone, in which the two long-branched taxa are not sister groups in the modeled (true) tree. In this case, parsimony retrieves an incorrect tree while likelihood succeeds in choosing the correct tree. (B) The Farris zone, in which the long-branched taxa are sister groups in the modeled tree. Parsimony considers the correct tree as the optimal solution while likelihood retrieves an incorrect tree (for datasets of up to 10,000 characters).

Huelsenbeck, 1993; Tateno *et al.*, 1994; Gaut and Lewis, 1995; Huelsenbeck, 1995; Nei *et al.*, 1995; Yang, 1996). As a result, parsimony was usually seen as the method suffering from a pervasive kind of flaw, long-branch attraction (LBA), in an area of the tree space, now called the Felsenstein zone.

A few simulation studies explored other kinds of 4T tree space, in which the long branches were sister taxa in the true topology (Fig. 1B; Russo *et al.*, 1996; Yang, 1996, 1997; Siddall, 1998; Bruno and Halpern, 1999). In datasets of up to 10,000 characters, likelihood performance was poor and, when analyzed, parsimony performed much better (Russo *et al.*, 1996; Yang, 1996; Siddall, 1998). The results found were opposite to those of the Felsenstein zone, and thus this segment of the tree space is now called the Farris zone (Siddall, 1998) or the anti-Felsenstein zone (Bruno and Halpern, 1999). Siddall (1998) proposed the presence of a bias in likelihood results, naming it long-branch repulsion (LBR), since the long branches were not recovered as sister taxa. However, Yang (1996) attributed these results to LBA in parsimony becoming an advantage of this method when the true tree happens to have the long branches clustered together. What is not clear from these simulation studies is whether parsimony outperforms likelihood for reasons other than merely succumbing to its own vices at the right time.

In the present approach the performance of parsimony and likelihood is analyzed in a more complex scenario, a 10-taxon case in which there are 4 long and 13 short branches. The choice of this restricted part of the potentially vast 10-taxon tree space was made in order to analyze the effect of the simultaneous presence of long branches being sister and nonsister taxa. Although this approach is not necessarily realistic, the complexity of a 10T case can reveal biases that affect real data analyses that cannot be revealed by a 4T case study.

MATERIALS AND METHODS

The datasets were created using a Markov model of character transformation and a defined topology over which the simulations were conducted. The parameter that was varied in these simulations was the length of the long branches, while the length of the short branches was kept constant. The resulting datasets were analyzed using parsimony and maximum likelihood. The results of each method were analyzed for the detection of different types of topological errors related to the increment of the long branch lengths.

Simulations

In the Markov model, the simulated datasets were produced with PAML version 2.0g (Yang, 1999). The stochastic model that determined the character transformation was HKY85 (Hasegawa *et al.*, 1985), with approximately equal base composition and a transition / transversion ratio of $\kappa = 5$. The α parameter of the Γ distribution was set to produce no substitution rate variation among sites. The topology chosen consists of a 10-taxon tree in which there are 4 long branches, arranged in two pairs of sister taxa (Fig. 2). The lengths of the short terminal and internodal branches were fixed at 0.01 and 0.02, respectively. The length of the long branches was increased from 0.1 to 1.5 with an increment of 0.1 between each set of replicates (Fig. 2); thus, long branches ranged from being 10 to 150 times the length of other terminal branches. One hundred dataset replicates of 1 kb were produced for each of the 15 long branch lengths. Additionally, two other sets of replicates were produced with longer sequence lengths (5 and 10 kb) in order to assess the behaviors for increasing amounts of similarly structured information.

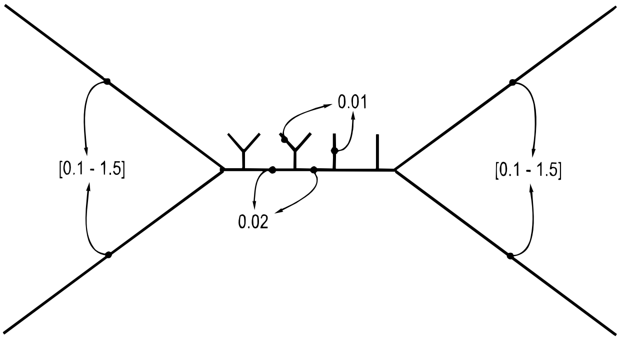

FIG. 2. The modeled tree of the 10T case analyzed in this work. There are 4 long branches, arranged in two pairs of sister taxa. The length of these long branches varied in the range indicated. The other 13 branches were modeled with the indicated fixed length.

Dataset Analysis

Parsimony. All datasets were analyzed in an equally weighted nonadditive parsimony analysis using PAUP* (Swofford, 1999). Additionally, the datasets were analyzed under a different character weighting scheme, with a transition/transversion cost equal to 5. Tree search strategy consisted of performing 100 duplicates of stepwise addition and TBR branch swapping.

Likelihood. The datasets of 1 kb were analyzed in PAUP* (Swofford, 1999) under two different stochastic models in the likelihood analyses. One of them was the true model (Hasegawa *et al.*, 1985), with the parameter values under which the sequences were created (nst = 2, variant = HKY, tratio = 2.5). The second model used was Jukes–Cantor (JC; Jukes and Cantor, 1969), the simplest model of stochastic evolution that uses a single parameter for all types of substitution (nst = 1). The larger datasets (5 and 10 kb) were analyzed only under the true model in order to briefly explore the effect of different sample sizes. Tree searches in maximum-likelihood analyses were performed with one round of TBR branch swapping applied to the best of the 14 different trees (including the true tree) that were found in all of the parsimony analyses of all the simulated datasets.

The estimation of branch lengths for the complete set of 1-kb matrices was analyzed with two different values of the maximum number of branch length smoothing passes (maxpass = 20 and maxpass = 1000). Since differences in the frequency of recovery of each topology were smaller than 1%, the analysis of larger datasets was done with 20 smoothing passes.

Bias Analysis

The analysis of results was focused on the detection of the most common topological errors. Here, these errors are referred to as biases when a systematic preference for an incorrect answer or a certain type of incorrect answer (i.e., types of errors) is detected in methods while analyzing finite datasets. These preferred incorrect answers are similarly referred to as biases in the literature of simulations (e.g., Bruno and Halpern, 1999). This terminology is used here too, although its use does not imply the strict statistical meaning of the term bias, since topologies are only awkwardly characterized as statistical parameters.

The diversity of incorrect topologies was analyzed and two different types of errors were recognized with respect to the location of the four long branches in accordance with their compatiblity with previously proposed biases (Fig. 3). A third type of error that does not involve the position of long branches also was counted (short-branch rearrangement) since it was observed among the results (Fig. 3). Since the aim of this study was the detection of tendencies to consider a particular kind of topology as optimal, the percentage of instances in which a method found the modeled tree and any of the three errors was measured in each set of replicates. These values were plotted against the length of the long branches in order to detect biases associated with that particular parameter variation

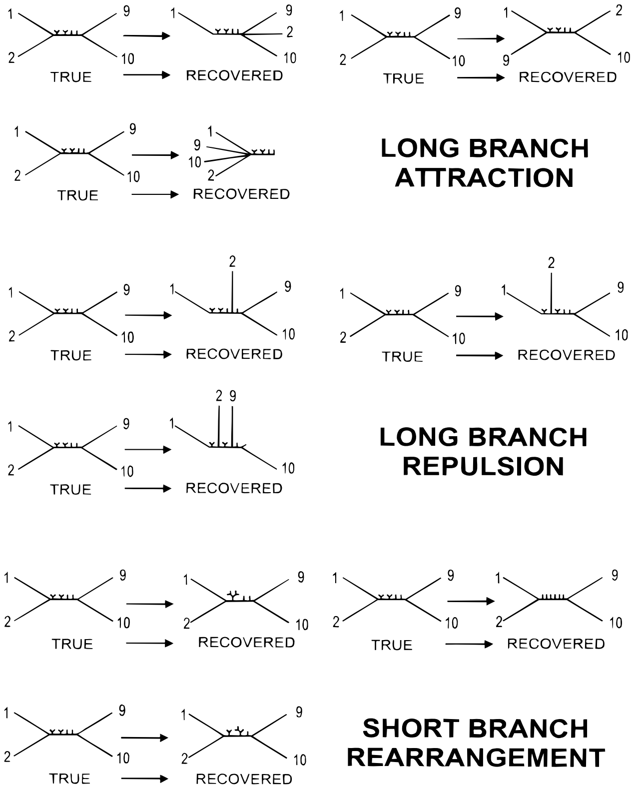

FIG. 3. Types of topological errors found in the present study with labeled long branches. Long-branch attraction topologies were counted when unrelated long branches were recovered as sister taxa in the optimal tree. Long-branch repulsion topologies were counted when at least two of the long branches were recovered as sister taxa of a short branch. The third type of error, short-branch rearrangement, was counted when an incorrect topology contained the correct pairs of long branches separated (clades 1+2 and 9+10).

(long branch length). In these plots the sum of all percentages can exceed 100% because of the presence of multiple optimal solutions for a particular replicate.

Kuhner and Felsenstein (1994) explored the accuracy of several methods in a simulation study in a 10T case. Although their approach to the 10T case was quite different from that analyzed here, they also were concerned with the existence of biases in phylogenetic methods. However, despite their interest in biases, the topological distance measures that they used for comparing the incorrect topologies are not sensitive to the distinction of the multiple biases that may exist in different methods (i.e., they would not specifically distinguish between LBR and LBA).

Character Sampling

As stated before, biases are properties of methods dealing with finite datasets, in contrast to statistical consistency, which is an asymptotic property of a method performing with an infinite sample of characters. However, the trends observed in successively larger but finite datasets were briefly explored in order to examine patterns within the range of dataset sizes that are normally used by practicing systematists. Thus, datasets with an increasing number of characters (1, 5, and 10 kb) also were analyzed with parsimony and with maximum likelihood assuming the correct model. This should not be interpreted as a prediction for statistical consistency because a seemingly monotonic trend observed in simulated finite datasets may be reversed in a larger analysis and thus does not ensure that the method is consistent (Kim, 1998; Siddall, 1998).

Taxon Sampling

The presence of long branches is commonly associated with biases in methods of phylogenetic reconstruction. More specifically, the interaction of two long branches commonly is the cautionary problem of phylogenetic methods. Since most of the observed errors in this study are related to the location of the long branches, the set of the simulated matrices of 1 kb, also was analyzed excluding some of those long branches, exploring the effect of the presence of one, two, or three long branches (Fig. 4). These modified matrices will be referred to as the 9T case (one long-branched taxon excluded), the 8T-Farris case (two long-branched sister taxa excluded), the 8T-Felsenstein case (two long-branched nonsister taxa excluded), the 7T case (only one long-branched taxon included), and the 6T case (all long-branched taxa excluded). These datasets were analyzed as above.

RESULTS

Perspectives on the performance of parsimony and likelihood in these simulations can vary. One might consider only the frequency of recovery of the "true"

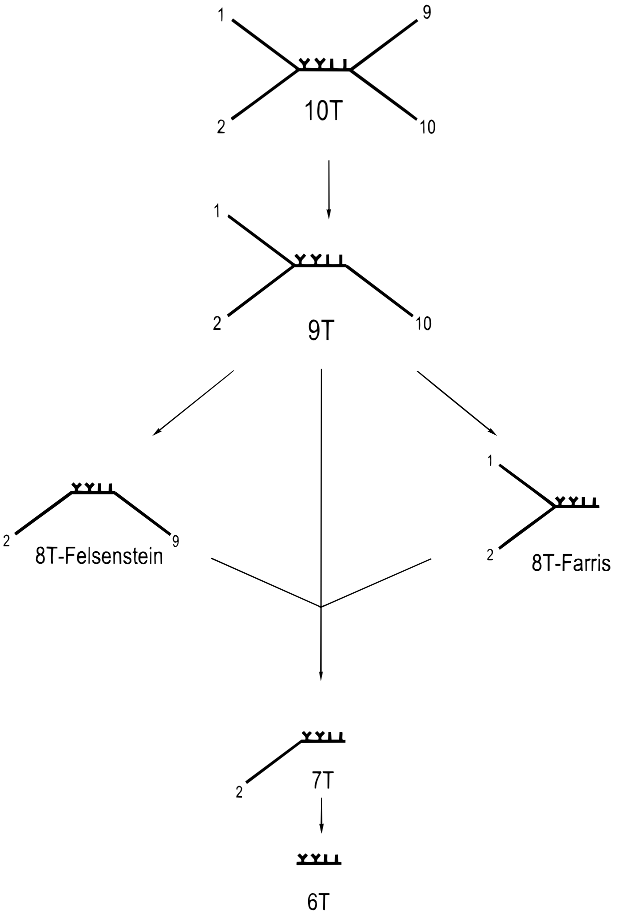

FIG. 4. Correct trees of the different taxon-sampling cases analyzed, when long branches were alternatively excluded from the datasets.

topology. However, as we have suggested, there is more than what is immediately apparent in such a consideration. There are the kinds of errors and biases found as well as how those errors are mitigated by increased information content. All three are considered here.

True Topology Recovery

Among the results both of parsimony and of likelihood analyses it was found that the performance of these methods decreases with increased long branch lengths. Two main differences are noted in this graph. First, differentially weighted parsimony (using a character weighting scheme similar to that of the simulated model) outperforms equally weighted parsimony, while in maximum likelihood the situation is the opposite. In accordance with previous results (Yang, 1997; Takahashi and Nei, 2000), it is seen here that maximum-likelihood analyses using a wrong model can outperform those using the correct model although the parsimony analysis using the correct model did better than equally weighted parsimony. Second, the parsimony and likelihood (true model) curves have different shapes (Fig. 5). Parsimony recovered the true topology with a high level of frequency when the length of the long branches was up to 0.4 or 0.5, depending on the character weighting scheme (i.e., up to 40 or 50 times the length of the other terminal branches). In this zone, parsimony recovered the modeled tree between 100 and 94% of the replicates, while likelihood's performance dropped almost linearly from 98% to 50%. Maximum likelihood (assuming the true model) outperformed parsimony when long branches were longer than 0.5 or 0.7 substitutions per site, depending on the character weighting scheme, but still performed poorly (Fig. 5). Parsimony, on the other hand, exhibited a logistic curve, decreasing only slowly at first and then rapidly losing performance through an inflection point slightly greater than 0.5 substitutions per site. For those who like direct extrapolations to real cases, it could be interesting to note that the zone in which parsimony has a high level of performance coincides with the

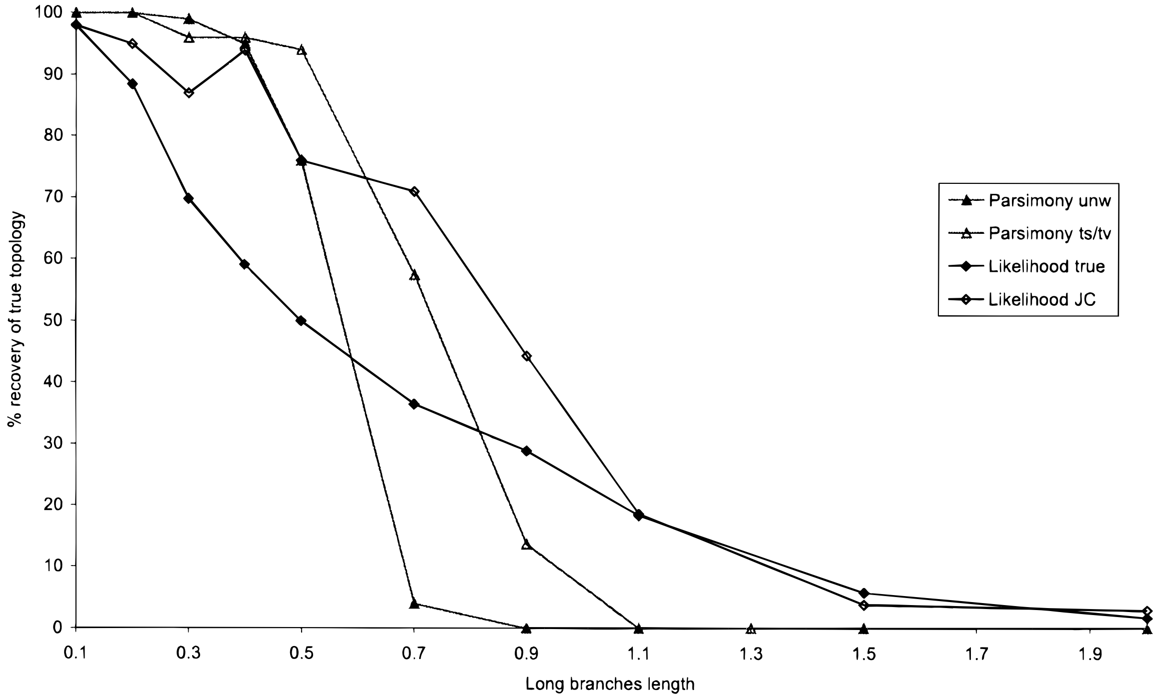

FIG. 5. Percentage of recovery of the correct topology in the 10T case (1-kb datasets) versus the length of the long branches. These curves correspond to equally weighted parsimony, differentially weighted parsimony, maximum likelihood assuming the true model, and maximum likelihood assuming the Jukes–Cantor model analyses.

branch length range inferred in the Biologically Meaningful Region known from real dataset analysis (Nei *et al.*, 1995).

Biases

As stated above, a principal interest of this work was not so much when methods get the right tree (Fig. 5) but how they get it wrong (Figs. 6–9).

As shown before, the performance of parsimony in the utilized region of the 10T tree space starts to decrease when long branch lengths are equal to or longer than 0.5 (i.e., 50 times the length of short branches). The incorrect topologies recovered in parsimony analyses almost invariably fall into only one of the recognized types of errors: long-branch attraction. Almost all errors had the four long branches clustered together but the pattern of relationships of the other short-branch taxa was wholly correctly recovered (Figs. 6 and 7). Only a low percentage of replicates found incorrect trees that do not fit into this category (less than 4% when long branch lengths varied between 0.3 and 0.7). Thus, the curves of recovery of the true topology and the LBA topology are essentially symmetrical in the two character weighting schemes explored (i.e., if parsimony retrieved a wrong answer, it was a topology compatible with LBA). These results closely fit the predictions made with previous simulation studies: if nucleotide transformations are driven by a Markov model of change (debatable in itself, but convincing to some), parsimony has a strong bias in preferring the clustering of long branches whether they are related or not. Since in a 10T case there are 2,027,024 wrong possible topologies, it is worth noting that parsimony systematically prefers only 15 of these when long branches are much longer than the other terminal branches. Parsimony suffers from a strong but identifiable bias (LBA).

The incorrect topologies recovered by likelihood analysis are rather different from those found by parsimony. In particular, there are several types of errors among the incorrect trees found by likelihood (Fig. 8). In the region of the tree space in which parsimony performs better (long terminal branches up to 50 times longer than the short terminal branches), incorrect topologies recovered by likelihood fall within two of the types of errors recognized. One of them, appearing

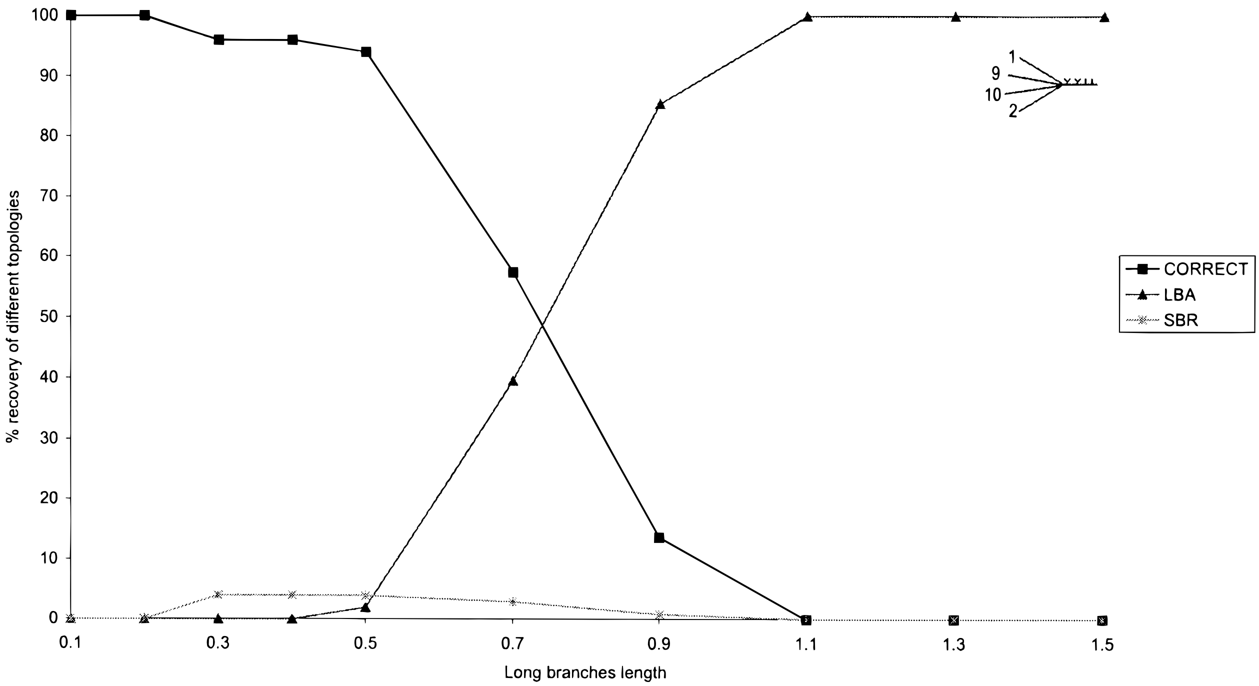

FIG. 6. Percentage of recovery of the correct tree and the different types of topological errors in the 10T case (1-kb datasets) versus the length of the long branches for the differentially weighted parsimony analyses. A representative of each kind of error is depicted beside each curve.

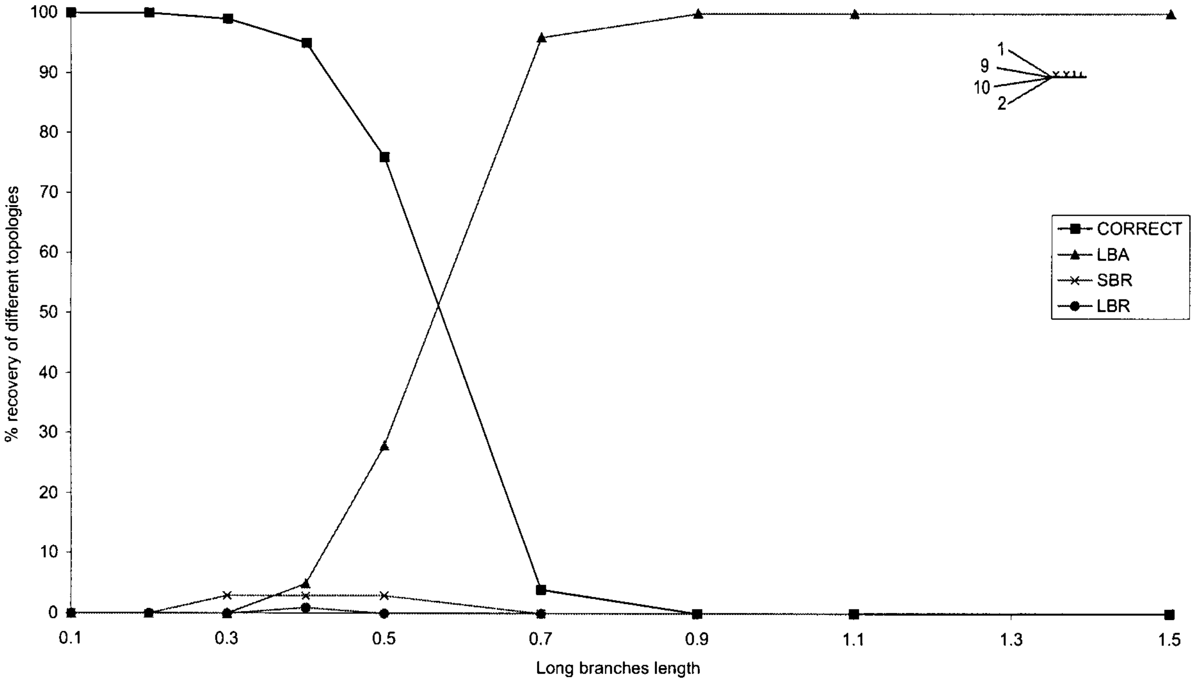

FIG. 7. Percentage of recovery of the correct tree and the different types of topological errors in the 10T case (1-kb datasets) versus the length of the long branches for the equally weighted parsimony analyses. A representative of each kind of error is depicted beside each curve.

when long branches were more than 10 times the length of the short terminal branches, was compatible with long-branch repulsion, a bias previously detailed by Siddall (1998). The other error found in this region appears among short branches and concerns considerable short-branch rearrangement (SBR). When the lengths of the long branches varied between 0.5 and 1.5 the proportion of LBR compatible topologies increased dramatically to 89%. In addition, other types of errors (LBA), start to appear with a frequency that varied between 1 and 16%.

When the long branch lengths were set to 2.0 substitutions per site, likelihood's performance was even worse. LBA compatible topologies increased slightly, reaching a frequency of recovery of 22% and LBR compatible topologies occurred in 86% of the errors.

The curve depicting recovery of the correct topology for likelihood with the simpler (JC) model is similar to that found with the true model, but JC performs better when long branches are lower than 1.0. The topological errors found by the JC likelihood analysis are similar to those with the true model, but differences exist in their frequencies (Fig. 9). There are only two errors found in likelihood analysis (JC) that steadily increase with the long branch length. LBA compatible topologies appears with shorter long branch lengths (0.7) but with higher frequencies (up to 43%) than when assuming the true model. Long branch attraction is already predicted to appear in likelihood analyses with oversimplified models (Yang, 1996; Bruno and Halpern, 1999). LBR compatible topologies appeared at the same long branch length values (0.1 substitutions per site) and reached higher levels of frequency of recovery (up to 43%) in the JC model analysis than was found using the correct model.

Character Sampling

The curves of true topology recovery in the equally weighted parsimony analysis have a similar shape for the three sets of matrices (1, 5, and 10 kb). Differences exist in the region in which long branch lengths are 40 to 60 times the length of other terminal branches (Fig. 10). In this region, parsimony behaved better as the number of characters increased. In other regions performance was observed to be equal for the different

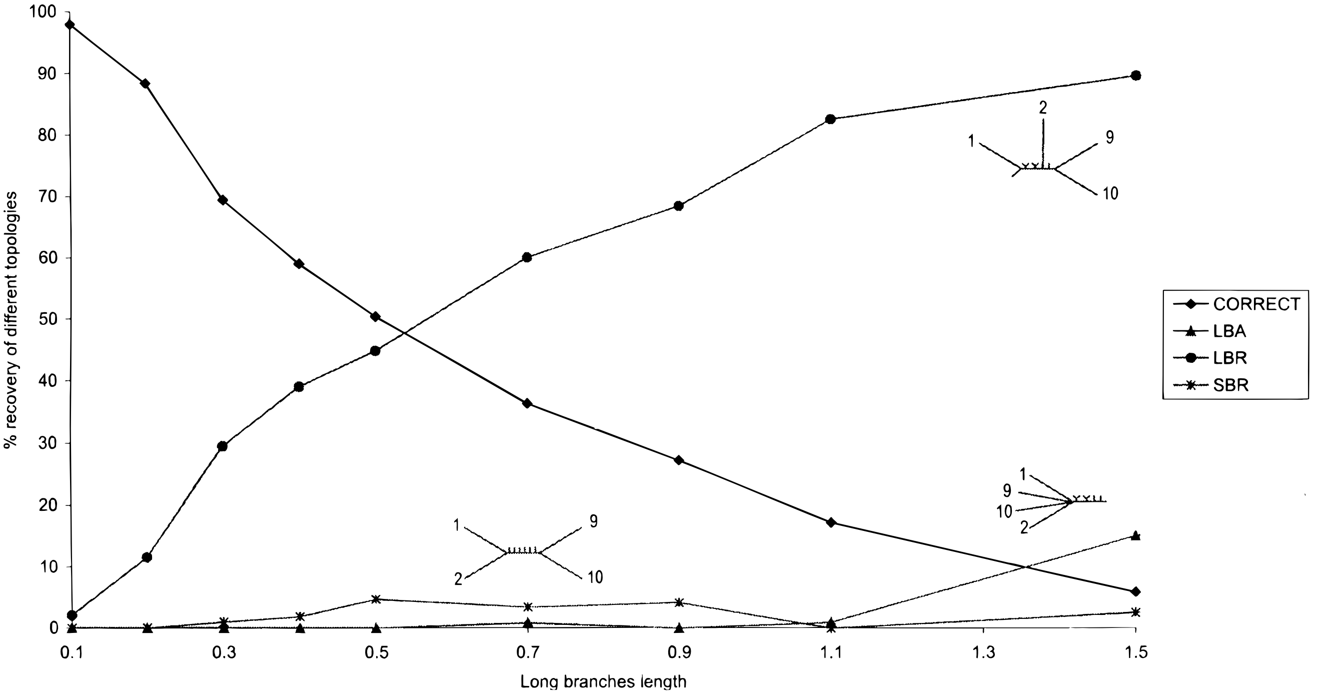

FIG. 8. Percentage of recovery of the correct tree and the different types of topological errors in the 10T case (1-kb datasets) versus the length of the long branches for likelihood analyses assuming the true model. A representative of each kind of error is depicted beside each curve.

sequence lengths (i.e., either 100% or 0%). All topological errors recovered in parsimony analyses of larger datasets (5 and 10 kb) fall within the LBA category.

In contrast to the pattern observed in parsimony, the likelihood (true model) analyses showed very different results in the larger datasets. Likelihood performance in recovering the true topology was better as the number of characters increased (Fig. 11). These results are not surprising since the likelihood analysis was done assuming the true model of character transformation and, thus, is a best-case scenario in which statistical consistency may be obtained. Nonetheless, maximum likelihood still performed worse than parsimony when the long branch lengths were equal to or lower than 0.5 substitutions per site.

Taxon Sampling

In the 6T case (all branches short) both parsimony and likelihood (irrespective of the model) recovered the true topology in 100% of the replicates. However, the results for the datasets containing any long-branched terminal taxa depicted different results for each method. In parsimony analyses, the addition of a single long-branched taxon (7T case) had little effect on the performance of the method in recovering the true tree (Figs. 12 and 13). When the long branch length was 1.5 substitutions per site, 17 or 24% of the topologies recovered were incorrect, depending on the character weighting scheme. The addition of two long branches resulted in different results depending on where the taxa were located in the true topology. Adding two long branches can have either drastic effects on the performance of the method (8T-Felsenstein) or almost none (8T-Farris). The curves of the 9T and the 10T cases show better performance than the 8T-Felsenstein case but a worse performance than the 8T-Farris case, as expected. Again, parsimony retrieves an incorrect topology only when two unrelated long branches are present in the tree.

The results of likelihood analyses with an increasing number of long branches are different from those for parsimony (Figs. 14 and 15). One of the most striking differences is that the presence of even a single long branch (7T case) produced a drastic drop in the ability to recover the true topology, irrespective of the model

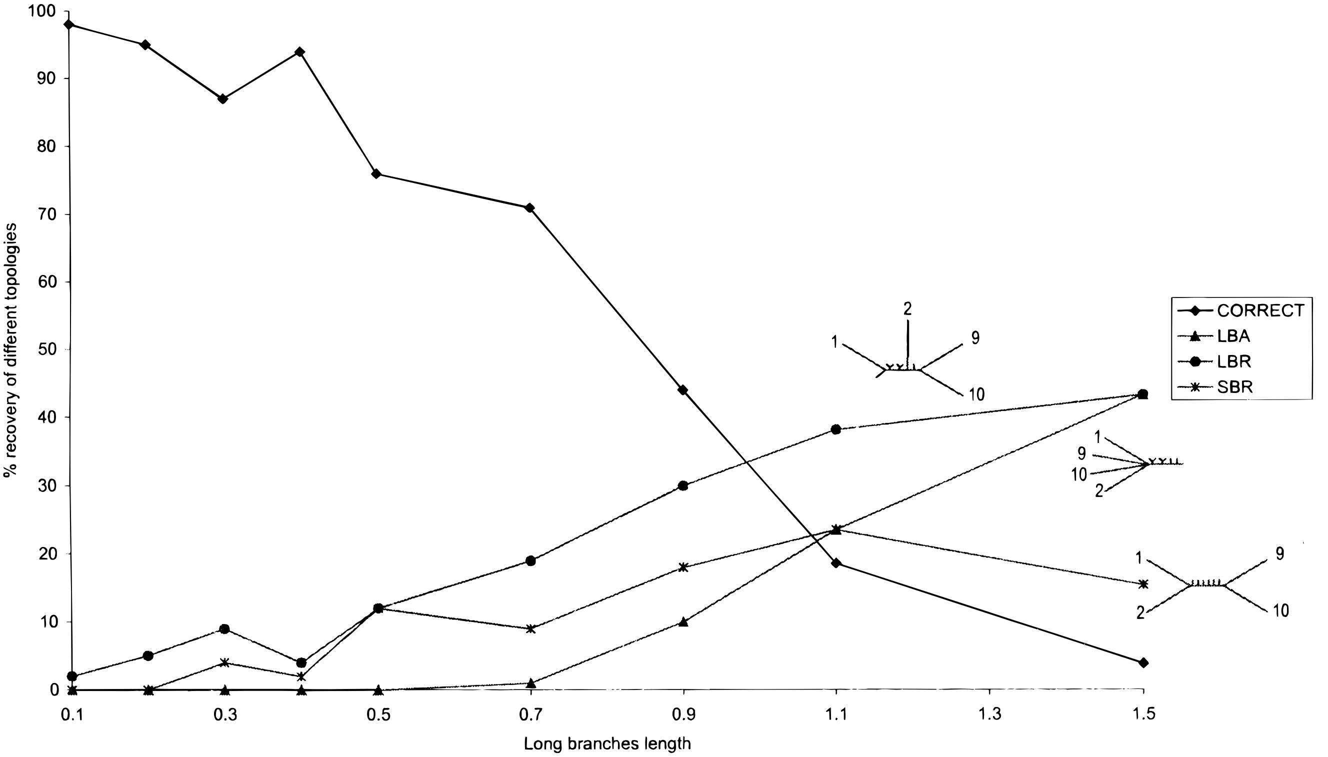

FIG. 9. Percentage of recovery of the correct tree and the different types of topological errors in the 10T case (1-kb datasets) versus the length of the long branches for likelihood analyses assuming the oversimplified JC model. A representative of each kind of error is depicted beside each curve.

(Figs. 14 and 15). The addition of two long branches affects this even more and occurs irrespective of whether they are related or unrelated taxa in the true tree (8T-Farris or 8T-Felsenstein). The 9T case performed similar to the 8T-Farris case under both models of analysis. The incorrect topologies recovered in these analyses showed the three recognized types of error (LBR, LBA, and SBR). There are some differences in the curves of topological errors recovered under the two assumed models, due mainly to the higher proportion of SBR errors recovered in the JC model analyses.

DISCUSSION

The aim of this study was to analyze the performance of maximum likelihood and parsimony in order to discover possible biases of these methods, focusing on a more complex scenario than previous 4T case studies (e.g., Huelsenbeck, 1995; Siddall, 1998). It could be useful to be aware of the existence of such biases, in order to be suspicious of the outcome of a particular method when the optimal tree is in a certain "zone" of the tree space. In several cases, phylogeneticists have questioned the results of real dataset parsimony analyses based on a suspicion of the presence only of the known LBA bias for this method (e.g., Allard and Miyamoto, 1992; Huelsenbeck, 1997, 1998; Stiller and Hall, 1999; Philippe and Germont, 2000; Sanderson *et al.*, 2000; Wiens and Hollingsworth, 2000). Similar circumspections were applied to likelihood results from some real dataset analyses, since it is well known that strong violations of the assumed model can lead to misleading results, such as LBA (Sullivan and Swofford, 1997).

An examination of the raw results of the parsimony and maximum-likelihood performances shows that both methods perform poorly when long branches coexist with short branches. Again, it should be noted that these results are not interpreted here as reasons

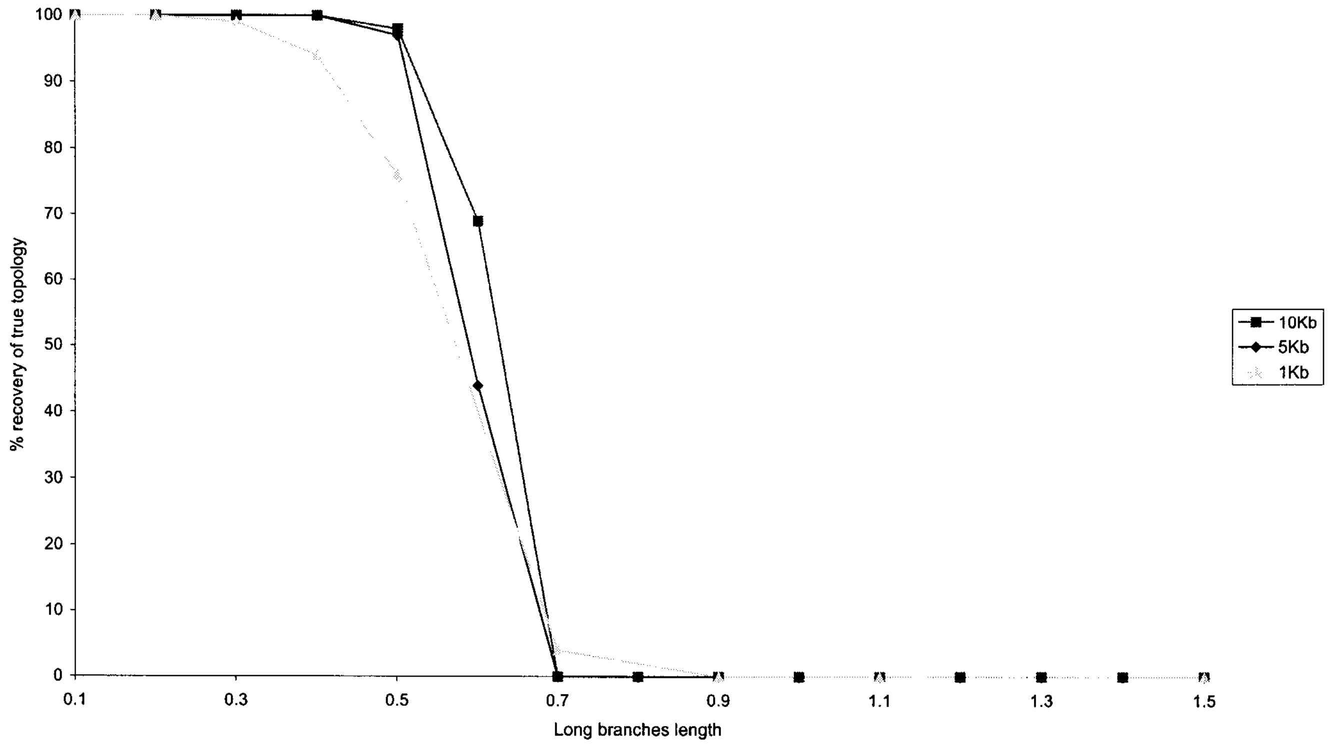

FIG. 10. Percentage of recovery of the correct topology in the 10T case versus the length of the long branches for the different character-sampling cases analyzed (1, 5, and 10 kb) in the equally weighted parsimony analyses.

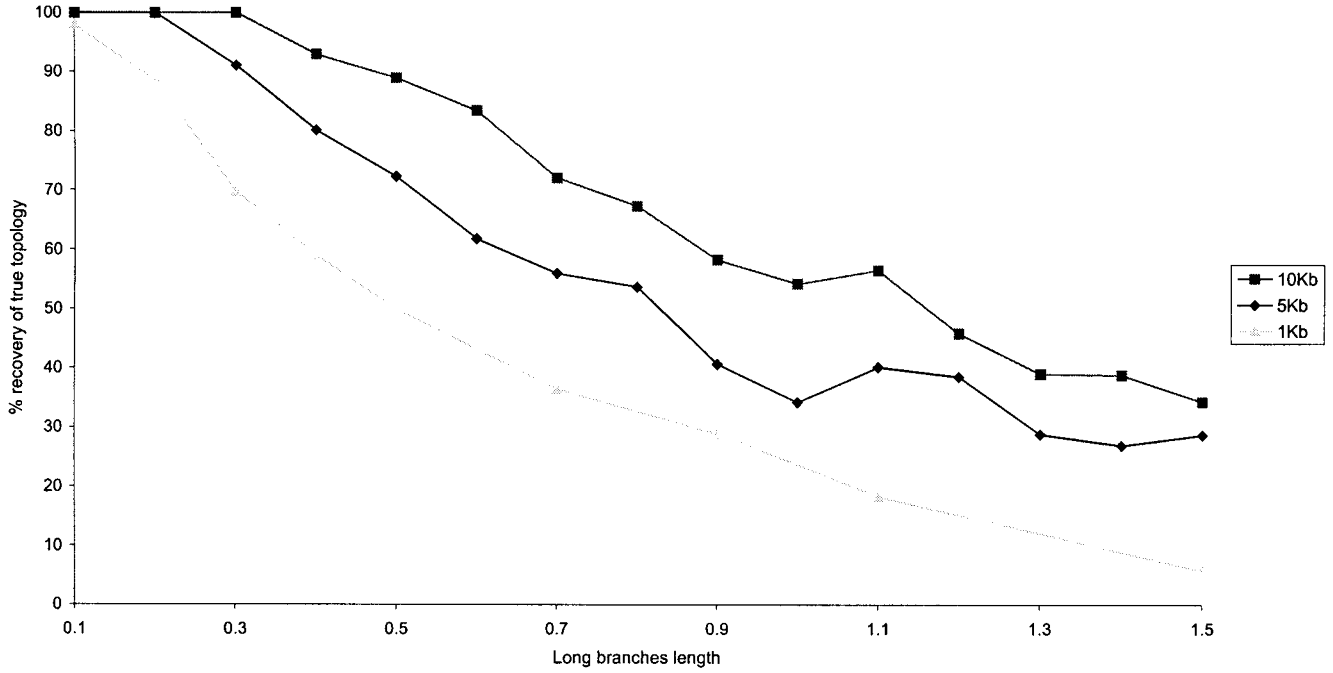

FIG. 11. Percentage of recovery of the correct topology in the 10T case versus the length of the long branches for the different character-sampling cases analyzed (1, 5, and 10 kb) in likelihood analyses assuming the true model.

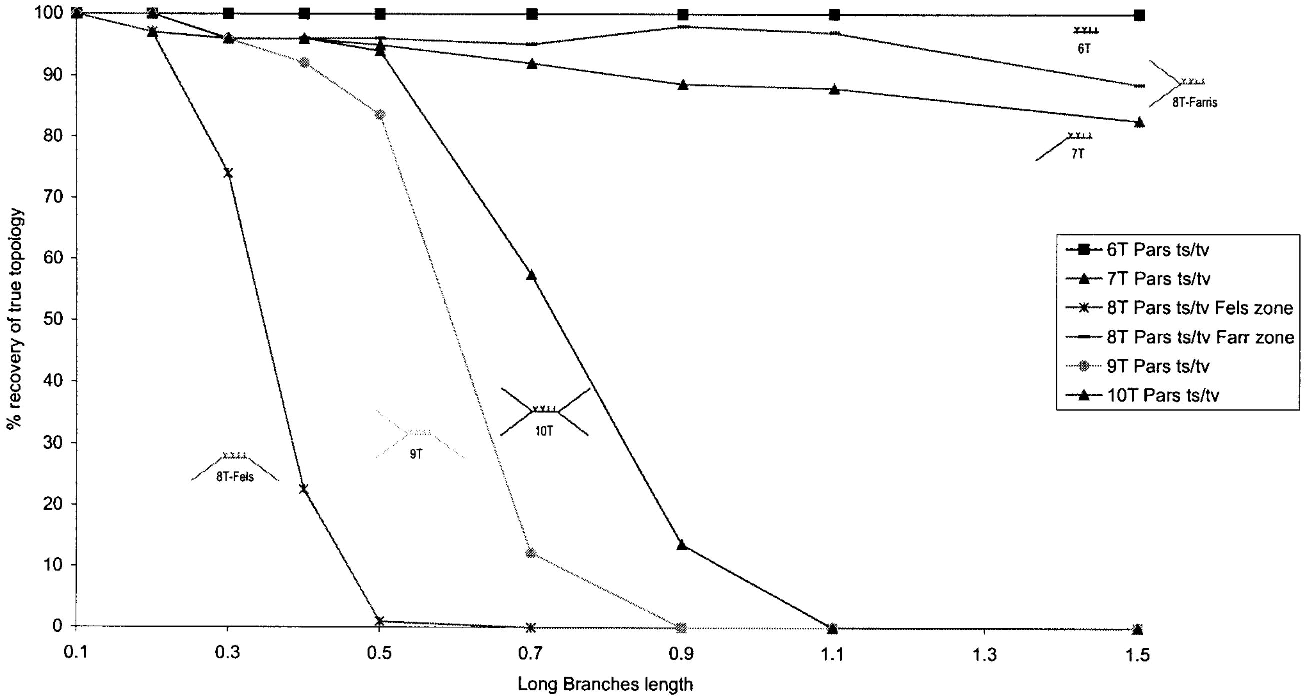

FIG. 12. Percentage of recovery of the correct topology versus the length of the long branches for the different taxon-sampling cases analyzed (1-kb datasets) in the differentially weighted parsimony analyses. The correct tree for each taxon-sampling case is depicted beside each curve.

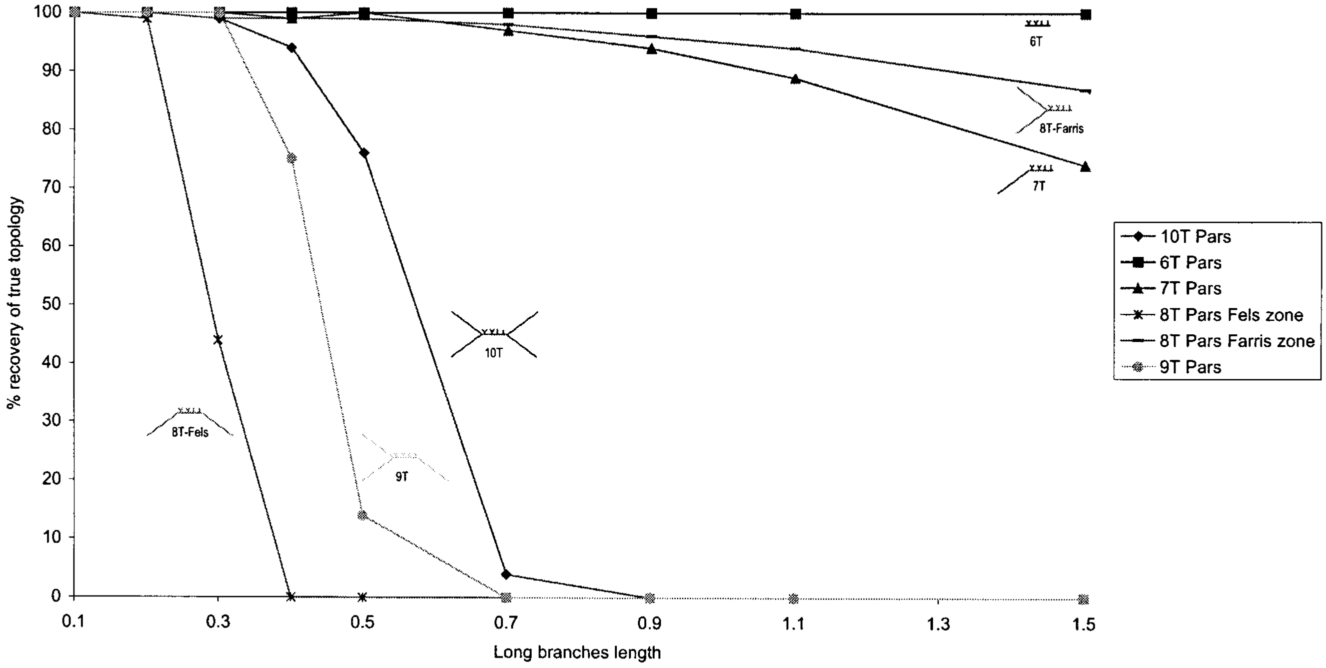

FIG. 13. Percentage of recovery of the correct topology versus the length of the long branches for the different taxon-sampling cases analyzed (1-kb datasets) in the equally weighted parsimony analyses. The correct tree for each taxon-sampling case is depicted beside each curve.

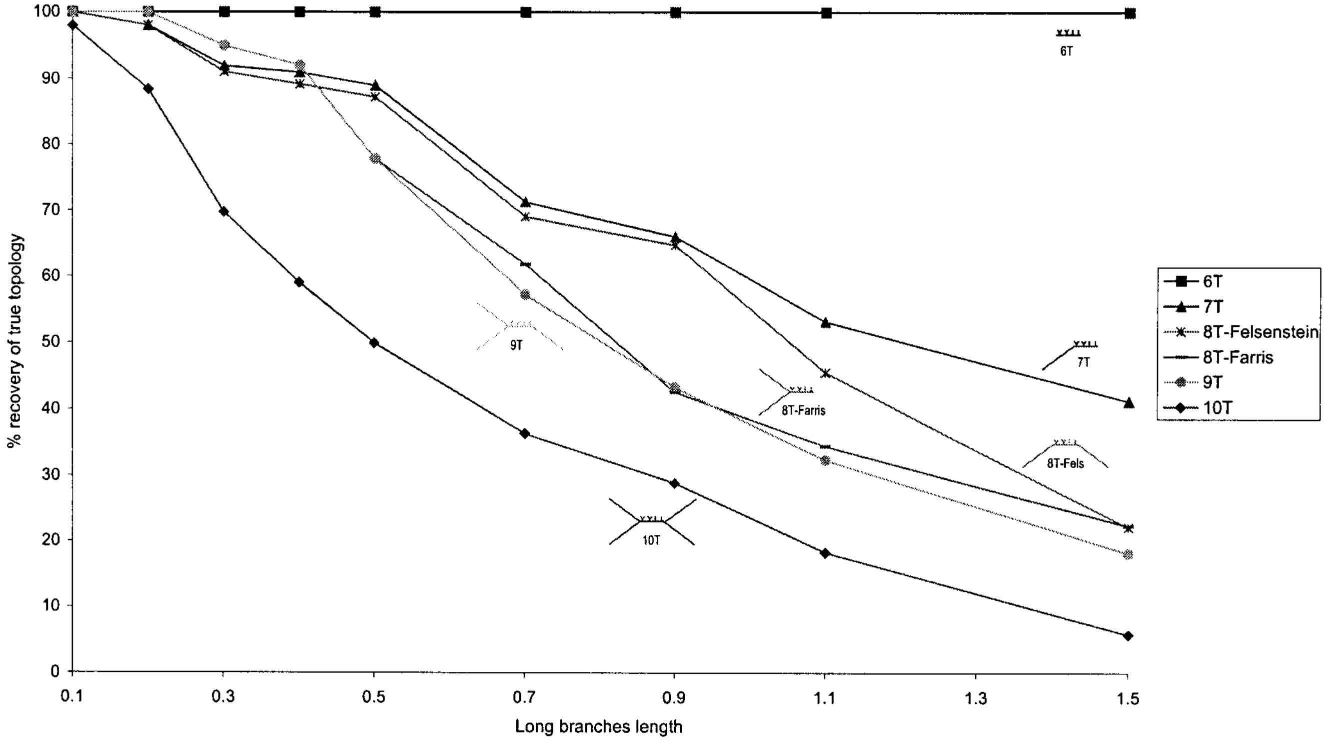

FIG. 14. Percentage of recovery of the correct topology versus the length of the long branches for the different taxon-sampling cases analyzed (1-kb datasets) in likelihood analyses assuming the true model. The correct tree for each taxon-sampling case is depicted beside each curve.

for discarding these methods, but rather as information concerning possible biases in these methods when heterogeneity in branch length is present. Although neither parsimony nor maximum likelihood properly handled long branches in these simulations, they do not suffer in the same way or from the same problems. The analyses of topological errors in the cases analyzed revealed the presence of only two strong bias for parsimony, long-branch attraction (as was also suggested by Kim (1993) in a 8T case). In contrast, in these cases, maximum likelihood did not infer the correct position of long branches retrieving topologies compatible with more than one kind of error (e.g., LBR, LBA, SBR), two of which were found to increase with the long branch length even when the model is correct. The presence of LBA in likelihood analysis with oversimplified models was previously noted, because such models would underestimate the actual number of substitutions in long branches, as parsimony does (Huelsenbeck, 1995; Gaut and Lewis, 1995; Yang, 1996, 1997; Sullivan and Swofford, 1997; Bruno and Halpern, 1999). However, the presence both of LBR compatible topologies under oversimplified models and of LBA topologies under the true model was unexpected.

The implication of these differences between parsimony and maximum likelihood is relevant for the identification of putative misleading results from analyses of real datasets. In such cases, the true topology is unknown, and knowledge of the existence of a bias is useful only if the misleading results are identified by a particular topological outcome. Long-branch attraction bias can be suspected if two long branches are depicted as sister groups in a parsimony analysis or in a likelihood analysis. In contrast, if two long branches are distant from each other, long-branch repulsion can be suspected only in a maximum-likelihood analysis, since parsimony does not suffer from this error in any case studied so far. If such bias-related topology is present, phylogeneticists can investigate the problem in order to determine what bias may be affecting the outcome of a particular analysis (Huelsenbeck, 1997, 1998; Siddall and Whiting, 1999).

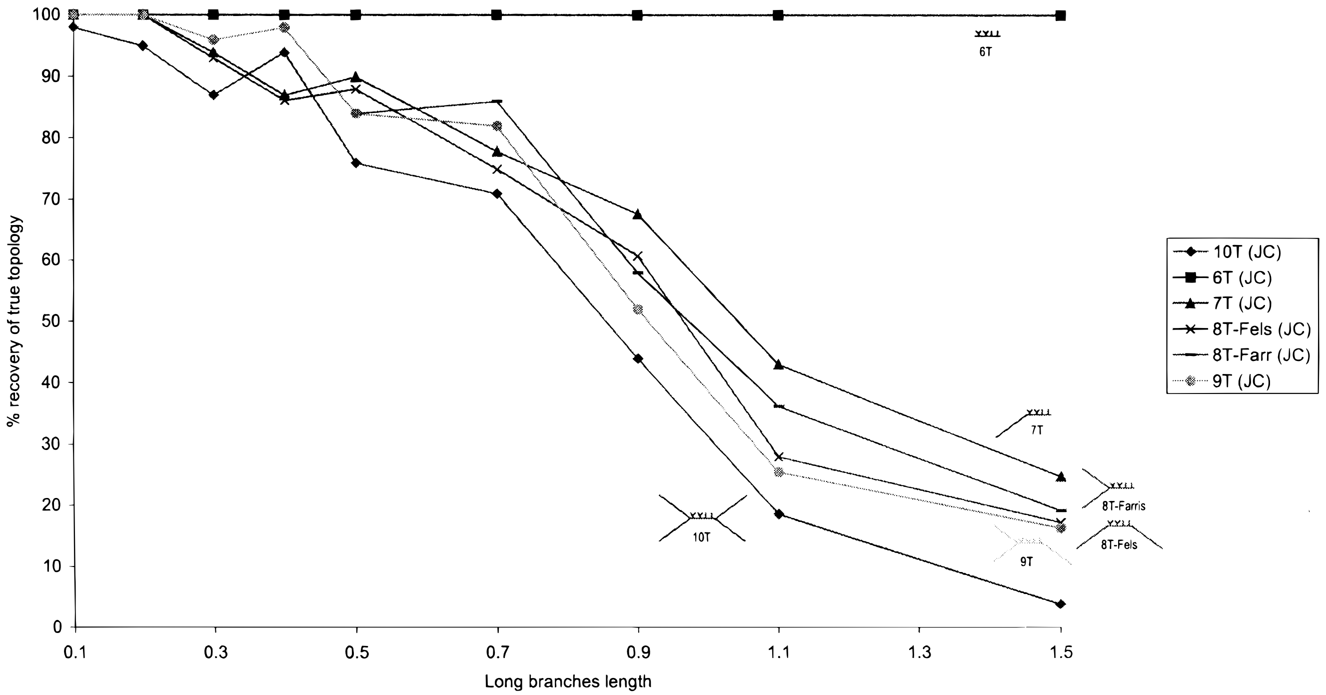

FIG. 15. Percentage of recovery of the correct topology versus the length of the long branches for the different taxon-sampling cases analyzed (1-kb datasets) in likelihood analyses assuming the oversimplified JC model. The correct tree for each taxon-sampling case is depicted beside each curve.

Parsimony suffers from only one bias (LBA) that is typified by a particular topological pattern (long branches depicted as sister taxa) making the identification of the putative presence of a misleading result easy in a real data analysis. If we consider parsimony as a valid method for phylogeny reconstruction, we can be suspicious of its results only if long branches are depicted as sister groups (since it is the only outcome of the only known bias).

In contrast, the existence of such a variety of topological errors in maximum-likelihood analyses makes the adoption of this strategy difficult. It appears from our results that the mere existence of marked branch length heterogeneity in a tree could be indicative of the putative existence of misleading results in a likelihood analysis, irrespective of their number or location.

An important difference between likelihood and parsimony relevant for the detection of long-branch-related errors in real dataset analysis is the influence of the presence of a single long branch in the modeled tree (7T case). The steep decrease in likelihood performance contrasts with the high level of performance observed in parsimony. Notably, this 7T case provides some information concerning the question of parsimony outperforming likelihood only because it happens to be biased in a useful direction. With only one long branch (7T), and no other to be "correctly" attached to, parsimony is behaving much like it does in the 8T-Farris in most of the explored tree space. Only when long branches are notably long, does the 8T-Farris outperform the 7T case in some replicates (by a maximum of 9 to 13 replicates, depending on the character weighting scheme), presumably due to LBA (Figs. 12 and 13). Thus, according to these results for extremely long branches, it is only in those 9 to 13% of replicates that parsimony retrieves the correct answer simply because it happens to be biased in a useful direction.

In contrast, likelihood is still suffering, probably from a failure to correctly estimate ancestral states. The most empirical way to evaluate a suspected LBA problem in a real dataset is to run the analyses while alternately excluding one of the long branches, since long branches cannot attract each other when they are not simultaneously present (Lyons-Weiler and Hoelzer, 1997; Siddall and Whiting, 1999). The results of the 7T case indicate that this strategy will be useful for parsimony analysis but may be misleading or uninterpretable in likelihood analysis. Other methods have been proposed for identifying LBA in real datasets (Hendy and Penny, 1993; Huelsenbeck, 1998; Sanderson *et al.*, 2000); however, these model-based approaches entail the assumptions that underlie the schism between supporters of cladistic and likelihood approaches and are unlikely to be adopted widely.

The behavior of datasets with longer sequences (5 and 10 kb) suggests that likelihood analysis can overcome its poor performance. However, this would be the case only if the whole sequence fits the assumed model. With increasing sequence lengths, it would be harder to justify the unique model assumption, and different segments of the sequence evolving under different models can lead to statistical inconsistency in maximum-likelihood analysis (Chang, 1996; Siddall, 1998; Farris, 1999).

REFERENCES

Allard, M. W., and Miyamoto, M. M. (1992). Testing phylogenetic approaches with empirical data, as illustrated with the parsimony method. *Mol. Biol. Evol.* **9,** 778–786.

Bruno, W. J., and Halpern, A. L. (1999). Topological bias and inconsistency of maximum likelihood using wrong models. *Mol. Biol. Evol.* **16,** 564–566.

Chang, J. T. (1996). Inconsistency of evolutionary tree topology reconstruction methods when substitution rates vary across characters. *Math. Biosci.* **134,** 189–215.

Edwards, A. W. F. (1995). Assessing molecular phylogenies. *Science* **267,** 253.

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 Univ. Press, New York.

Farris, J. S. (1999). Likelihood and inconsistency. *Cladistics* **15,** 199–204.

Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27,** 401–410.

Gaut, B. S., and Lewis, P. O. (1995). Success of maximum likelihood phylogeny inference in the four-taxon case. *Mol. Biol. Evol.* **12,** 152–162.

Hasegawa, M., Kishino, H., and Yano, T. (1985). Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22,** 160–174.

Hendy, M. D., and Penny, D. (1993). Spectral analysis of phylogenetic data. *J. Classif.* **10,** 5–24.

Hillis, D. M., and Huelsenbeck, J. P. (1993). Success of phylogenetic methods in the four-taxon case. *Syst. Biol. 42,* 247–264.

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. (1998). Systematic bias in phylogenetic analysis: Is the Strepsiptera problem solved? *Syst. Biol.* **47,** 69–74.

Huelsenbeck, J. P., and Hillis, D. M. (1993). Success of phylogenetic methods in the four-taxon case. *Syst. Biol.* **42,** 247–264.

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.

Kim, J. (1993). Improving the accuracy of phylogenetic estimation by combining different methods. *Syst. Biol.* **42,** 331–340.

Kim, J. (1998). Large-scale phylogenies and measuring the performance of phylogenetic estimators. *Syst. Biol.* **47,** 43–60.

Kuhner, M. K., and Felsenstein, J. (1994). A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol. Biol. Evol.* **11,** 459–468.

Lehmann, E. L. (1983). Theory of point estimation. Wiley, New York.

Lyons-Weiler, J., and Hoelzer, G. A. (1997). Escaping from the Felsenstein zone by detecting long branches in phylogenetic data. *Mol. Phylogenet. Evol.* **8,** 375–384.

Lyons-Weiler, J., and Takahashi, K. (1999). Branch length heterogeneity leads to nonindependent branch length estimates and can decrease the efficiency of methods of phylogenetic inference. *J. Mol. Evol.* **49,** 392–405.

Nei, M., Takezaki, N., and Sitnikova, T. (1995). Assessing molecular phylogenies. *Science* **267,** 253–255.

Philippe, H., and Germont, A. (2000). Phylogeny of eukaryotes based on ribosomal RNA: Long-branch attraction and models of sequence evolution. *Mol. Biol. Evol.* **17,** 830–834.

Russo, C. A. M., Takezaki, N., and Nei, M. (1996). Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. *Mol. Biol. Evol.* **13,** 525–536.

Sanderson, M. J., Wojciechowski, M. F., Hu, J.-M., Sher Khan, T., and Brady, S. G. (2000). Error, bias, and long-branch attraction in data for two chloroplast photosystem genes in seed plants. *Mol. Biol. Evol.* **17,** 782–797.

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 Whitting, M. (1999). Long-branch abstractions. *Cladistics* **15,** 9–24.

Stiller, J. W., and Hall, B. D. (1999). Long-branch attraction and the rDNA model of early eukaryotic evolution. *Mol. Biol. Evol.* **16,** 1270–1279.

Sullivan, J., and Swofford, D. L. (1997). Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. *J. Mammal. Evol.* **4,** 77–86.

Swofford, D. L. (1999). Phylogenetic Analysis using Parsimony (and other methods), PAUP v.4.0b2. Sinauer, Sunderland, MA.

Takahashi, K., and Nei, M. (2000). Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. *Mol. Biol. Evol.* **17,** 1251–1258.

Tateno, Y., Takezaki, N., and Nei, M. (1994). Relative efficiencies of the maximum likelihood, neighbor-joining, and maximum parsimony methods when substitution rate varies with site. *Mol. Biol. Evol.* **11,** 261–277

Wiens, J. J., and Hollingsworth, B. D. (2000). War of the iguanas: Conflicting molecular and morphological phylogenies and long-branch attraction in iguanid lizards. *Syst. Biol.* **49,** 143–159.

Yang, Z. (1996). Phylogenetic analysis using parsimony and likelihood methods. *J. Mol. Evol.* **42,** 294–307.

Yang, Z. (1997). How often do wrong models produce better phylogenies? *Mol. Biol. Evol.* **14,** 105–108.

Yang, Z. (1999). PAML version 2.0g, Phylogenetic Analysis by Maximum Likelihood. Program and documentation (http://abacus.gene.ucl.ac.uk/software/paml.html).