Syst. Biol. 50(4):525–539, 2001

Bias in Phylogenetic Estimation and Its Relevance to the Choice between Parsimony and Likelihood Methods

DAVID L. SWOFFORD,[1,6] PETER J. WADDELL,[2] JOHN P. HUELSENBECK,[3]
PETER G. FOSTER,[1,7] PAUL O. LEWIS,[4] AND JAMES S. ROGERS[5]

[1]*Laboratory of Molecular Systematics, National Museum of Natural History, Smithsonian Institution Museum Support Center, 4210 Silver Hill Road, Suitland, Maryland 20746, USA; E-mail: swofford@lms.si.edu*
[2]*Institute of Molecular BioSciences, Massey University, Palmerston North, New Zealand; E-mail: waddell@onyx.si.edu*
[3]*Department of Biology, University of Rochester, Rochester, New York 14627, USA; E-mail: johnh@brahms.biology.rochester.edu*
[4]*Department of Ecology and Evolutionary Biology, The University of Connecticut, U-43, 75 N. Eagleville Road, Storrs, Connecticut 06269-30437, USA; E-mail: plewis@uconnvm.uconn.edu*
[5]*Department of Biological Sciences, University of New Orleans, New Orleans, Louisiana 70148, USA; E-mail: jsrogers@uno.edu*

It is now widely recognized that under relatively simple models of stochastic change, phylogenetic inference methods can actively mislead investigators attempting to estimate evolutionary trees from molecular sequences and other data. One instance of this phenomenon is "long-branch attraction," in which some pairs of taxa have a higher probability of sharing the same character state because of parallel or convergent changes along long branches than do taxa that are more closely related because they have retained some same state from a common ancestor. Methods that systematically underestimate the actual amount of divergence may then become statistically inconsistent or "positively misleading" (Felsenstein, 1978; Hendy and Penny, 1989), estimating an incorrect tree with an increasing certainty as the amount of character data increases. Although usually associated with parsimony methods, long-branch attraction can also afflict maximum likelihood and distance analyses when the assumed substitution models of these methods are strongly violated (e.g., Huelsenbeck and Hillis, 1993; Huelsenbeck, 1995; Waddell, 1995:377–404; Gaut and Lewis, 1995; Chang, 1996a; Lockhart et al., 1996; Sullivan and Swofford, 1997). In this case, although the methods are explicitly designed to deal with superimposed substitutions (multiple hits), the underlying models predict fewer of these than actually occur and thus do not go far enough in correcting for the problem. Inconsistency can also arise under parsimony, even when all branches have the same length (Kim, 1996), although in this case there must still be particular imbalances in the total lengths of the paths from internal nodes to tips of the tree; "long-path attraction" would describe this phenomenon.

Long-branch attraction has been widely used, and abused, in justifying choices of methods and in explaining anomalous results. Critics of the relevance of long-branch attraction and related artifacts have generally taken two tacks. The first (e.g., Farris, 1983) claims that the demonstration of long-branch attraction requires simple and unrealistic models of evolutionary change. As pointed out by Kim (1996), this argument lacks force because conditions that lead to inconsistency are much more general and complex than those outlined by Felsenstein (1978); further relaxation of Felsenstein's conditions simply exacerbates the problem. The second line of argument (e.g., Siddall and Kluge, 1997) follows from the fact that "truth" is unknowable in science generally; because it is not possible to be certain that the analysis of a real data set has been compromised by long-branch attraction, the ability of a method to converge, in principle, to the correct solution with increasing amounts of data is irrelevant. In this view, "'accuracy' is rendered empty as an empirical claim" (Siddall and Kluge, 1997:318). Proponents of model-based (or statistical) methods that seek to avoid inconsistency attributable to long-branch or long-path artifacts have not been dissuaded by this argument. They certainly appreciate

[6]Current address: The Natural History Museum, Cromwell Road, London SW7 5BD, U.K.; E-mail: p.foster@nhm.ac.uk

[7]Current address: School of Computational Science and Information Technology, Florida State University, Tallahassee, Florida, 32306-4120.

the elusiveness of "truth" but understand that all methods are susceptible to failure under certain conditions. Consequently, these proponents seek methods and models that will succeed under a wide range of plausible conditions and that are less likely to yield misleading results purely because of artifacts. Historically, the different perspectives have led to a schism between those who would approach phylogenetics from a statistical perspective and those who place strong faith in one particular approach over all others. In many areas of science, the statistical modeling viewpoint tends to become more predominant as a subject matures. However, proponents of model-based methods in phylogenetics have not always helped their case by making overly assertive and sometimes misleading claims about the superiority of these methods (see Sidow, 1994; Hillis et al., 1994).

Against this backdrop of confusing and often acrimonious debate, Siddall (1998) offered a new challenge to the position that considerations of long-branch attraction favor model-based methods. Siddall's position, which seems reasonable at least on the surface, can be summarized simply: Although maximum likelihood and corrected-distance methods outperform parsimony methods in the so-called Felsenstein zone (four-taxon tree with two long, but unrelated, terminal branches and all other branches short), parsimony is better able to infer the correct tree topology in what Siddall calls the Farris zone, where the two long terminal branches instead lead to sister taxa (or are adjacent on an unrooted tree). Thus, if an unrooted phylogeny contains two long (terminal) branches plus three short branches, and the long branches are expected to lead to sister taxa about as often as they lead to nonsister taxa, then one might argue that there is no compelling reason for preferring one method over another on the basis of long-branch attraction. Waddell (1995) too had earlier referred to the Farris zone, calling it the "anti-Felsenstein" zone. Neither of these designations seems entirely appropriate, and we will use the term "inverse-Felsenstein zone" here. Siddall refers to the poor performance of maximum likelihood in the inverse-Felsenstein zone as "long-branch repulsion," a term used by Waddell (1995) for the significantly different problem of performance in the inverse-Felsenstein zone when the model is misspecified (e.g., overcorrection for among-site rate variation). However, we will use this term in Siddall's context for the present purposes.

We—and undoubtedly others—realized long ago that when long-branch attraction favors the correct unrooted tree for four taxa rather than one of the two incorrect trees, parsimony would outperform maximum likelihood *in choosing a topology*. Parsimony "succeeds" in the inverse-Felsenstein zone because it is a strongly biased method, the direction of the bias favoring the correct tree rather than an incorrect one, in contrast to the situation in the Felsenstein zone. This point was obvious enough not to merit publication on its own, although we have mentioned it in various other contexts (e.g., Swofford et al., 1995; Waddell, 1995). However, the portrayal by Siddall (1998) of this observation as a victory for parsimony methods demands closer scrutiny. Properly interpreted, the results of Siddall's simulations actually support the superiority of model-based methods for dealing with long-branch artifacts just as strongly as did those from earlier studies that concentrated on the Felsenstein zone. We emphasize at the outset, however, that the following analysis is not intended as a general criticism of the parsimony method. Rather, we show that results such as those of Siddall (1998) should not be taken as a vindication of parsimony with respect to one particular problem—sensitivity to long-branch-attraction artifacts.

Performance of Maximum Likelihood in the Inverse-Felsenstein Zone

Siddall's (1998) simulation results are summarized in Figure 1. Siddall's branch-length parameters were defined (Siddall, 1998:212) as the "expected percentage change of the . . . branches." This refers to the expected percentage of sites for which the nucleotide at one end of a branch (internode or edge) differs from the nucleotide at the other end. To avoid ambiguity, we prefer to call this quantity the expected percentage difference. Under the model used for his simulations, this value, expressed as a proportion p, is a lower bound on the expected number of changes (substitutions) per site including multiple hits, which we will call d. The two measures are related by using the familiar distance

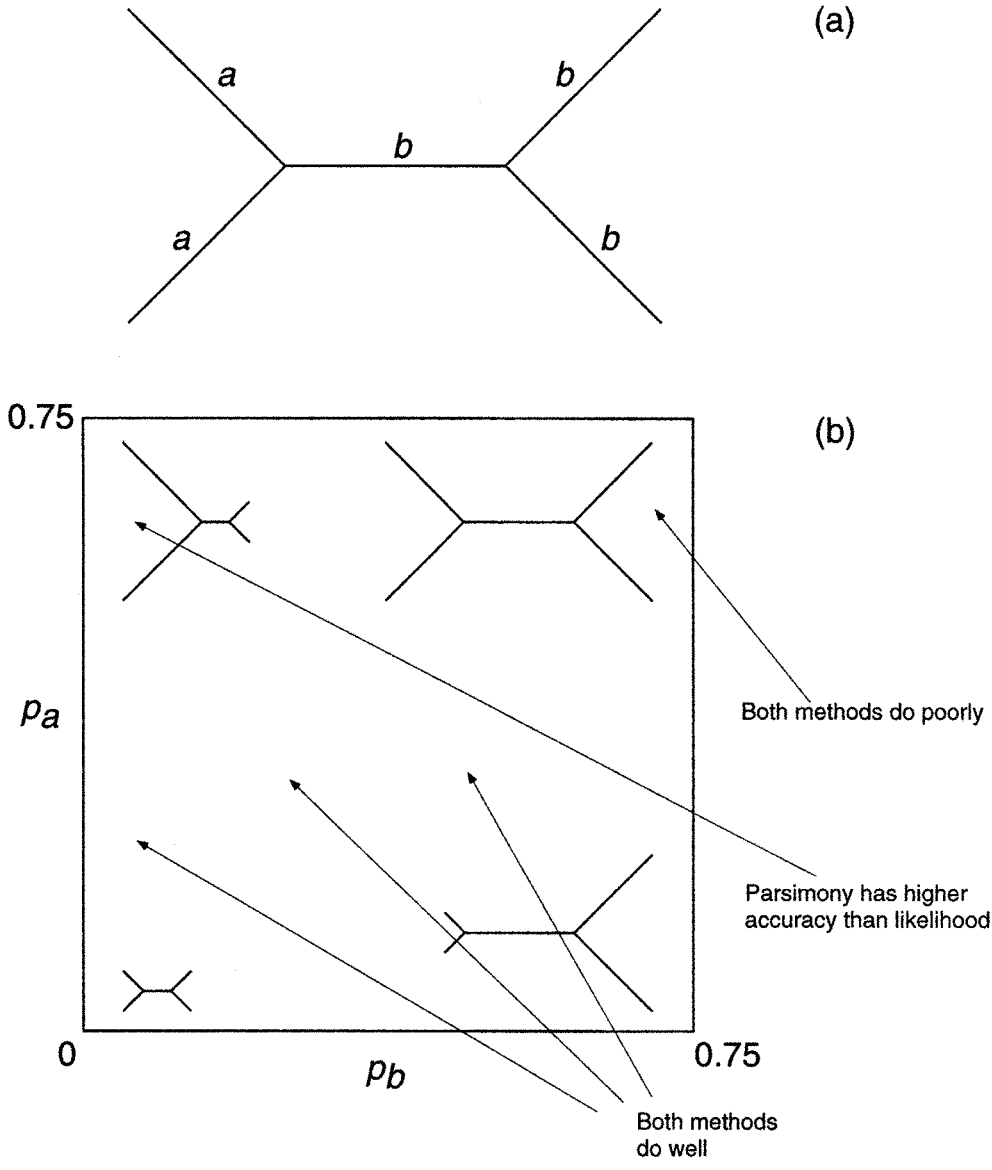

FIGURE 1. (a) Four-taxon model tree used by Siddall (1998). The probability of a difference in character states between the nodes incident to branches labeled a and b is given by p_a and p_b, respectively. (b) Parameter-space investigated by Siddall, showing relative performance of parsimony and likelihood (under the Jukes–Cantor model) in various regions of this space.

equation of Jukes and Cantor (1969):

$$d = -\frac{3}{4}\ln\left(1 - \frac{4}{3}p\right)$$

and its inverse:

$$p = \frac{3}{4} - \frac{3}{4}\exp\left(-\frac{4}{3}d\right).$$

Thus, the longest branch length simulated by Siddall, $p = 0.75$, corresponds to an infinitely long branch, and the next longest, $p = 0.675$, corresponds to a mean of about 1.7 substitutions per site along the branch.

Close examination of Siddall's (1998) simulation results immediately reveals some anomalies. The first involves his claim (1998:213 and his Fig. 4) that parsimony achieved high accuracy when all branch

lengths were $p = 0.75$, in which case each sequence was fully randomized with respect to all others. A method that could successfully reconstruct the true topology most of the time from completely random data would be a powerful one indeed, but parsimony is not this method. We repeated these simulations using a research version of PAUP*4.0d65 written by the first author. (With the long, biologically implausible, branches on trees simulated here, the likelihood surface with respect to branch-length parameters becomes extremely flat, so convergence of these parameters to their optimal values is very slow. To adjust for this, the limit on the maximum number of passes over the tree, *MaxPass*, was increased from the default value of 20 to 1,000, thus minimizing the possibility that failure of likelihood to converge to an optimal solution might affect accuracy rates.) Our results, shown in Table 1, are in complete accord with the prediction that given random data, parsimony can do no better than picking a tree at random, with a 1 in 3 chance of choosing the correct tree. For long but finite branch lengths ($p = 0.675$), parsimony performs somewhat better than a random tree selection, but even for 1,000 sites parsimony has a <50% chance of correctly inferring the tree. Thus, Siddall's statement that "with 1000 characters free to vary, [parsimony] reconstructed the correct model tree more than 95% of the time across the whole parameter" space is clearly untrue. Note that with long but finite branch lengths, maximum likelihood slightly outperforms parsimony (Table 1, $p = 0.675$ columns). This result is also at variance with Siddall's (1998:213) statement that "likelihood methods also recovered the correct topology with somewhat lower accuracies than parsimony when all branch rates were equal but high." Siddall claimed that this "same phenomenon" was evident in Huelsenbeck's (1995) simulation study but was "not noted." However, reexamination of Huelsenbeck's Appendix 1, where the relevant comparisons are presented (Huelsenbeck, 1995:37, rows 1 and 5), reveals no qualitative difference in the relative performance of parsimony and likelihood in the upper right corner of the graphs.

A second anomaly in Siddall's (1998) presentation is the suggestion that in the inverse-Felsenstein zone, the accuracy of likelihood methods declines irreversibly with increasing sequence length: "... 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" (Siddall, 1998:213). This result is in direct opposition to theoretical predictions. Chang (1996b) and Rogers (1997) have independently proved that on binary trees with finite branch lengths, maximum likelihood is guaranteed to be statistically consistent when characters evolve according to a common mechanism under the assumptions of the model. These proofs establish that when assumptions of the model are met, as they are in these simulations, maximum likelihood methods should converge toward 100% accuracy with increasing sequence length at any point in the inverse-Felsenstein zone (or any other zone), except for points involving

TABLE 1. Performance of parsimony and likelihood (under the Jukes–Cantor model) when all five branch lengths are equal and long.

Number of sites	Method	Branch lengths			
		$p = 0.75$		$p = 0.675$	
		Prop. correct[a]	Prop. correct[b]	Prop. correct[a]	Prop. correct[b]
100	Parsimony	0.403	0.3375	0.446	0.3770
	Likelihood	0.477	0.3395	0.513	0.3808
500	Parsimony	0.352	0.3205	0.472	0.4368
	Likelihood	0.426	0.3478	0.511	0.4453
1000	Parsimony	0.363	0.3405	0.487	0.4632
	Likelihood	0.389	0.3258	0.530	0.4828

[a]Proportion of correctly estimated trees in 1,000 simulation replicates using Siddall's system for handling tied trees. When more than one optimal tree is found, the result is considered fully correct if the true tree is contained in this set.

[b]Proportion of correctly estimated trees in 1,000 simulation replicates using our preferred system for handling tied trees. One-half credit is given if the true tree is one of two optimal trees, one-third credit is given if all three trees have equal scores.

infinite-length branches. Our own simulations are in accord with this prediction (as was acknowledged in Siddall's "note added in proof"), although the success rate does not monotonically approach perfect accuracy. For example, Figure 2 shows the results of our simulations for one fairly extreme inverse-Felsenstein-zone point evaluated by Siddall (1998). The accuracy of likelihood is higher for 100 sites than for 500 or 1,000 sites, so in the absence of relevant theory, one could not fault an investigator for guessing that the accuracy of the likelihood method might continue to decline with still longer sequences. However, the relevant theory does exist, and consistent with its prediction, increasing sequence length enables likelihood to eventually turn the corner and begin moving gradually toward 100% accuracy. In this example, the phylogenetic problem is simply so difficult that it is unreasonable to expect any relatively unbiased method to perform well without an extremely large amount of data, and the simulations confirm this intuition.

Rather than considering the possibility of an error in his simulations, Siddall (1998) adopts the position that Felsenstein's (1978) claim for the consistency of maximum likelihood estimation of phylogenetic trees, based on earlier work of Wald (1949), is not valid. Siddall (1998:215), following Farris (1997, 1999) and possibly Yang (1996), asserts that

> 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.

Neither part of this statement is true. In principle, the sites of a nucleotide distribution certainly *can* be independently and identically distributed, whether or not they actually are so distributed in any particular case. Siddall and Kluge's (1997) earlier assertion that nucleotide characters cannot logically be independent is based on a

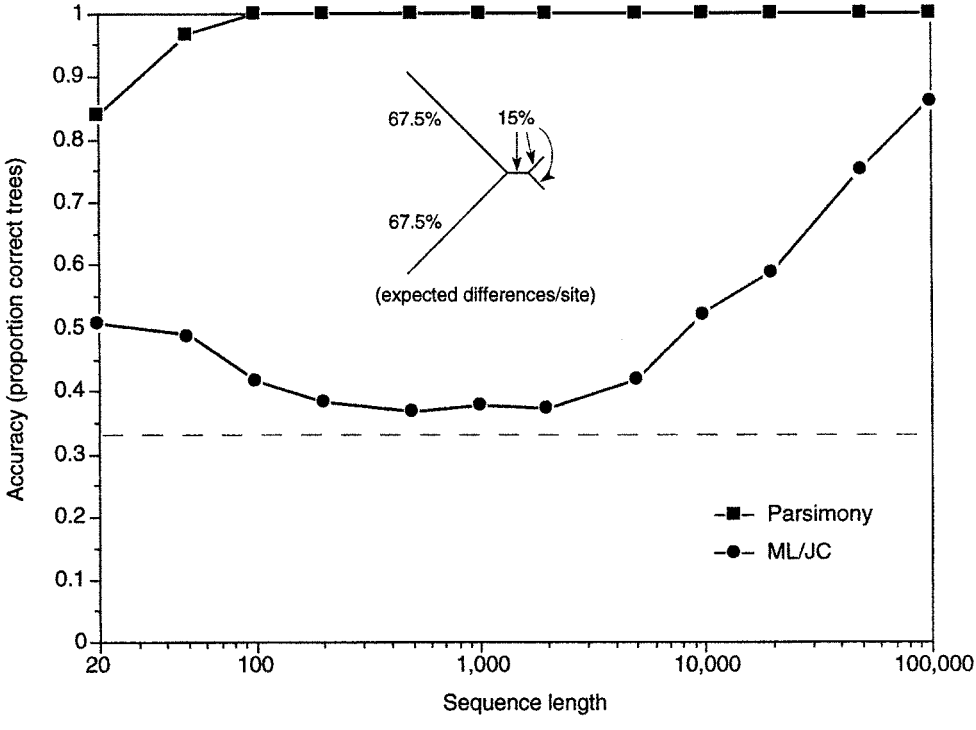

FIGURE 2. Performance of parsimony and likelihood with increasing sequence length for one point in the inverse-Felsenstein zone ($p_a = 0.675$, $p_b = 0.15$). Accuracy is measured as the proportion of correctly estimated trees in 1,000 simulation replicates. Parsimony achieves nearly perfect accuracy with only 50 sites. The accuracy of likelihood with very short sequences is helped by a bias favoring the correct tree. As sequence length increases, the bias exerts less influence and accuracy initially declines, but eventually moves toward the predicted 100% accuracy.

fundamental misunderstanding of the nature and application of the independence assumption. In any case, correlation between sites does not preclude consistency as long as the strength of the correlation decays at some very minor rate (Waddell et al., 1997). With regard to the claim of a requirement of continuity and differentiability of the likelihood function, Wald (1949:595) states explicitly that his proof "make[s] no differentiability assumptions (thus, not even the existence of the likelihood equation is postulated)." Furthermore, Chang (1996b) explicitly treats tree topology as a parameter in his proof of the consistency of maximum likelihood for estimating trees, which he refers to as a "customized variant" of Wald's proof. For $L(\theta)$, the likelihood function with respect to a parameter θ, the "likelihood equation" referred to by Wald (1949) is

$$\frac{\partial L(\theta)}{\partial \theta} = 0,$$

(Kendall and Stuart, 1979:39). If such an equation exists, the optimal value of θ can be solved for as one of the roots, either explicitly or by iterative procedures such as Newton's method. This simplifies finding the optimal value of θ, but it is not a requirement for the consistency of the maximum likelihood method. In the case of an unordered, discrete-valued parameter such as tree topology, this simply means we must use some other method for searching the parameter space (e.g., a tree-searching algorithm such as exhaustive search, branch-and-bound, or branch swapping) to attempt to find the optimal "value" of this parameter. Although clumsier and more time-consuming than conventional mathematical solution procedures, this requirement is merely a practical problem for the method, not a theoretical one.

Confusion over this latter point may arise from Wald's declaration of his Assumption 1— "$F(x, \theta)$ is either discrete for all θ or is absolutely continuous for all θ." As Wald explains in the preceding sentence, $F(x, \theta)$ is the cumulative distribution function of the random variable x, which in the case of sequence data is the nucleotide site pattern, a discrete variable. So, in this case the distribution of the random variable is discrete and the second part of Assumption 1 does not apply. However, even in the case of continuous distribution functions, the required absolute continuity is with regard to the random variable x, not to the parameters θ. In the continuous case, continuity is required so that the probability density function $f(x, \theta)$, which is related to the distribution function by the equation

$$f(x, \theta) = \frac{\partial F(x, \theta)}{\partial x},$$

will always exist for all values of x. In the discrete case, as Wald notes, $f(x, \theta)$ is the probability of x, not the probability density, so the requirement of differentiability does not arise at all.

Bias in Maximum Likelihood Estimation

An estimator is biased if its expected value differs from its true (population) value. Even when maximum likelihood is consistent, it is not guaranteed to be unbiased. A well-known example is the maximum likelihood estimator of a population variance when the data are drawn from a normal distribution,

$$s^2 = \sum (X - \bar{X})^2 / n,$$

where n must be replaced by $n - 1$ to obtain an unbiased estimator. (In this case, X is a continuous random variable, but see Kuhner and Felsenstein (1994) for one approach to quantifying bias on discrete tree topologies.) When two terminal branches on a four-taxon tree are extremely long and the remaining three branches are short, maximum likelihood tree inference under the Jukes–Cantor model is affected by bias. The presence of bias is suggested by the results shown in Figure 2, where the performance of likelihood declines initially and then improves as sequence length increases. In this case, although estimates of underlying parameters (branch lengths) are biased, maximum likelihood manages to obtain a correct tree topology more often than either of the incorrect topologies at all sequence lengths. This is not always the case. Figure 3 shows that in the Felsenstein zone where the two long branches are not adjacent on the tree, a bias in likelihood causes the

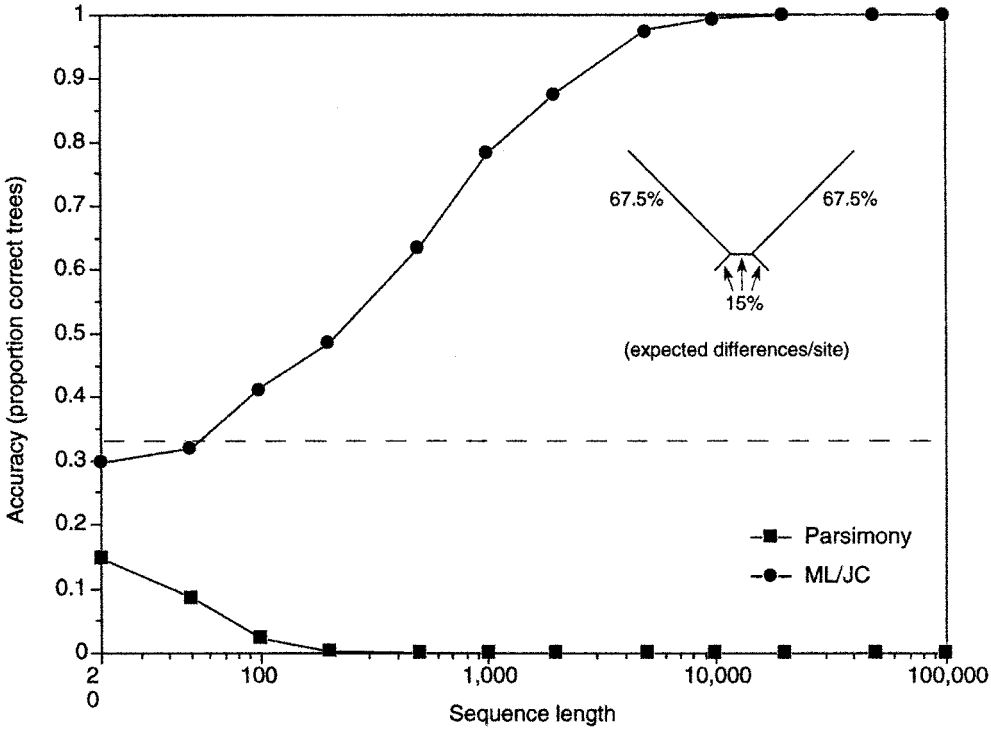

FIGURE 3. Performance of parsimony and likelihood with increasing sequence length for the point in the Felsenstein zone analogous to the one used for the inverse-Felsenstein zone simulations of Figure 2. Accuracy is measured as the proportion of correctly estimated trees in 1,000 simulation replicates. At 50 or fewer sites, likelihood actually does slightly worse than picking a tree at random (because of bias), but with increasing sequence length the bias decays and the correct tree is recovered with increasing certainty.

accuracy rate for likelihood for very short sequences to be lower than randomly picking a tree. However, as the consistency proofs guarantee, the bias is eventually overcome and the accuracy of likelihood increases toward 100% with longer sequences. The only conditions under which Siddall's conclusion of equal preference for all three possible trees is realized involve infinitely long branches (whereas the consistency proofs require finite branch lengths). Because infinite branch lengths are not reasonable biologically, the performance of a method under these conditions is not highly relevant to the choice between methods, although one would hope that a method would not return a strong preference for any one four-taxon tree when two of the four sequences are completely random (see below).

BIAS IN PARSIMONY ANALYSIS

Despite the errors in Siddall's simulation results and their interpretation, his primary conclusion is correct—in the inverse-Felsenstein zone of four-taxon branch-length space, parsimony estimates a phylogeny correctly more often than does maximum likelihood. We think it informative to examine the reason for the superior performance by parsimony under these conditions.

For the simple model of evolution simulated by Siddall, one can calculate the probability that an apparent synapomorphy uniting the two long-branch taxa is in fact due to homoplasy. (Here, we use the term synapomorphy in an unrooted sense; it will correspond to its traditional meaning if any one of the four terminal taxa is designated as an outgroup.) In the extreme end of the inverse-Felsenstein zone, an overwhelming number of apparent synapomorphies link the two long-branch taxa together. However, the synapomorphies uniting the two long branches can arise in many different ways. Figure 4 illustrates a few of the different character histories that can lead to an apparent synapomorphy linking the long-branch taxa. For this example, the nucleotides observed at

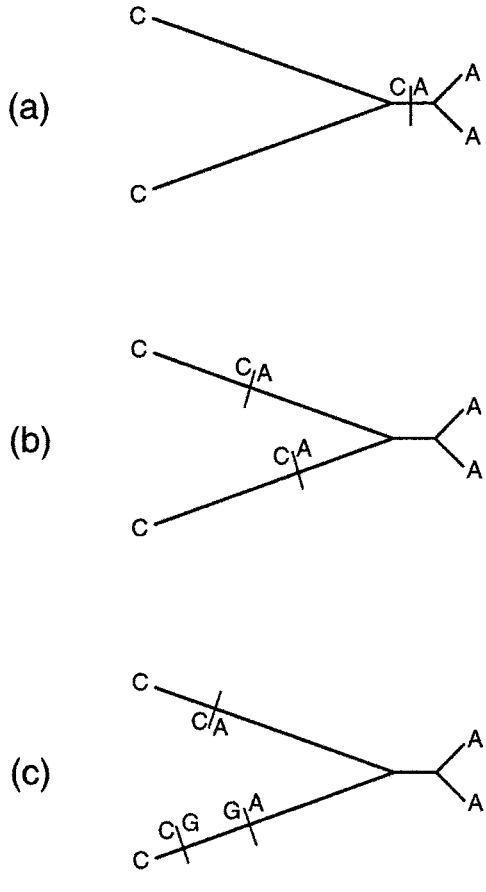

FIGURE 4. Different scenarios that will lead to an apparent synapomorphy. (a) A "true" synapomorphy. (b,c) Two scenarios in which an apparent synapomorphy is actually the result of misinterpreted homoplasy.

the tips of the tree are C for the long-branch taxa and A for the remaining two taxa. All of the examples except that shown in Figure 4a involve more than one change. In all cases, however, the parsimony method interprets the history of the character as a single change that occurred along the internal branch of the tree. In other words, except for the single example of Figure 4a, parsimony misinterprets homoplasy as evidence of relationship (in this case, as a relationship uniting the two long-branch taxa). This would not be a problem for the parsimony method if the probability is small that homoplasy underlies the apparent synapomorphies. In the inverse-Felsenstein zone, however, a vast majority of the apparent synapomorphies uniting the long-branch taxa are due to homoplasy. Consider the point in the parameter space analyzed in Figure 2, where the expected numbers of changes on long and short branches are 1.727 and 0.167, respectively. The probability that a single change will occur along the internal branch but no change will occur along the remaining branches for this tree is

$$\begin{aligned}
&\Pr[\text{True Synapomorphy}] \\
&\quad = \Pr[\text{No change on long terminal branches}] \\
&\qquad \times \Pr[\text{No change on short terminal branches}] \\
&\qquad \times \Pr[\text{single change on internal branch}] \\
&\quad \approx (e^{-1.727})^2 (e^{-0.167})^2 (0.167 e^{-0.167}) \\
&\quad \approx 0.0032
\end{aligned}$$

On the other hand, the probability of observing a pattern of nucleotides in which the two taxa on one side of the central branch share the same nucleotide and that nucleotide is different from a nucleotide shared by the two taxa on the opposite side of the central branch is 0.1172 (obtained as the sum of the single-site likelihoods for all $xxyy$-type patterns, e.g., AACC, AAGG, . . ., TTCC). Thus, about $(0.1172 - 0.0032)/0.1172$, or 97%, of all apparent synapomorphies will actually be misinterpreted homoplasies. At more extreme points of the parameter space examined by Siddall, the misinterpretation becomes even more pronounced. For example, at the second most extreme point simulated by Siddall ($p_a = 0.675$, $p_b = 0.0075$), ~99.8% of apparent synapomorphies supporting the true tree will in fact be misinterpreted homoplasies!

Figures 5 and 6 summarize the relative contribution of actual synapomorphies (those apparent synapomorphies that arise from a single change along the internal branch of the tree) versus misinterpreted homoplasy for the full parameter space explored by Siddall (1998). Figure 5a shows the expected proportion of parsimony-informative sites for which the two internal nodes have a different state and each pair of adjacent taxa have the same state (this includes both true synapomorphies as well as sites for which multiple substitutions have occurred along a branch but parsimony correctly reconstructs the ancestral states). The expected proportion of parsimony-informative sites that are apparent synapomorphies resulting from homoplasy in the long-branch taxa is shown

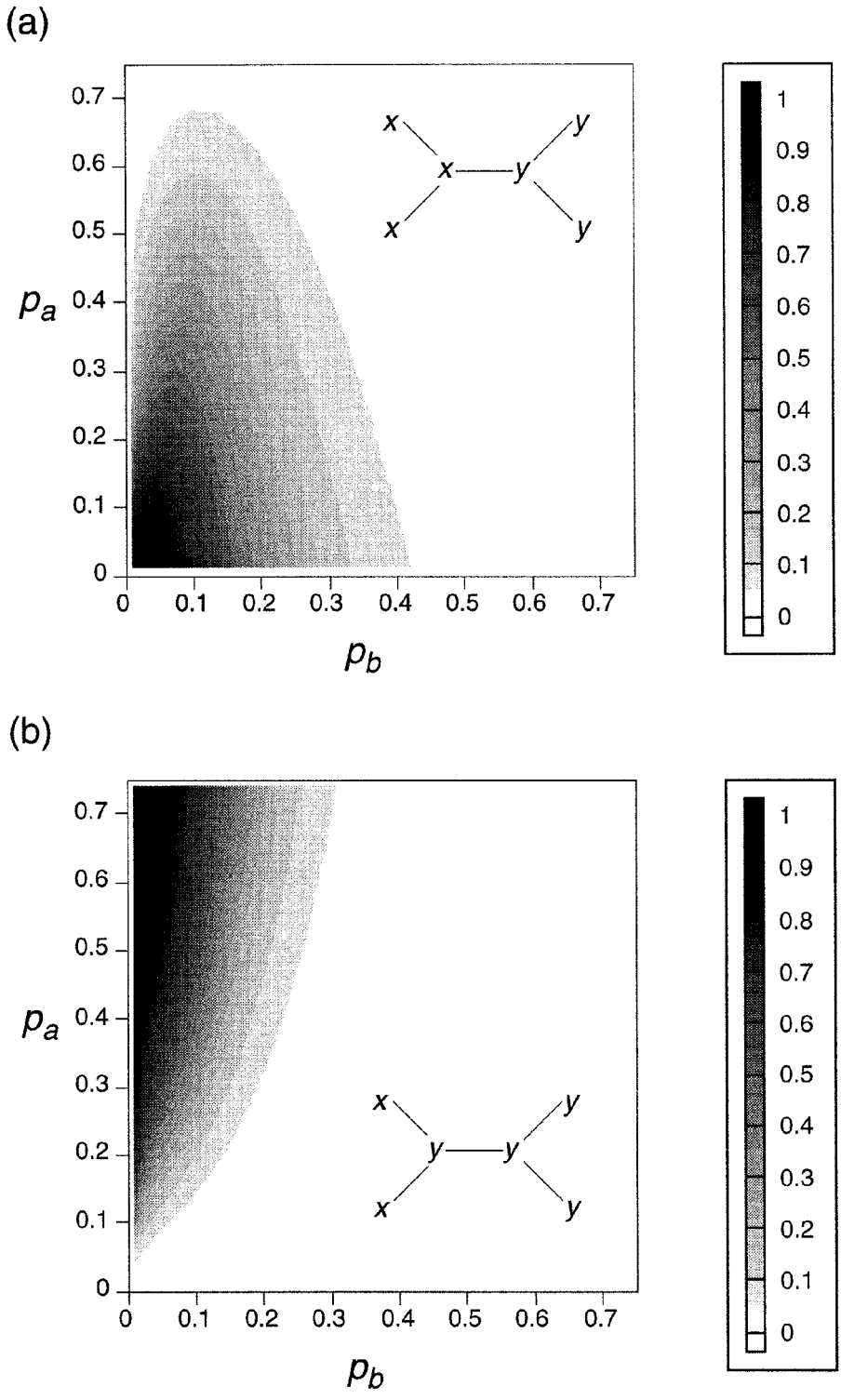

FIGURE 5. Contour plots showing the proportion of parsimony-informative sites for which (a) parsimony correctly reconstructs the states at the internal nodes, and this reconstruction suggests an apparent synapomorphy, and (b) parsimony misinterprets parallel changes in the terminal branches as a synapomorphy. See Figure 1 for definition of p_a and p_b.

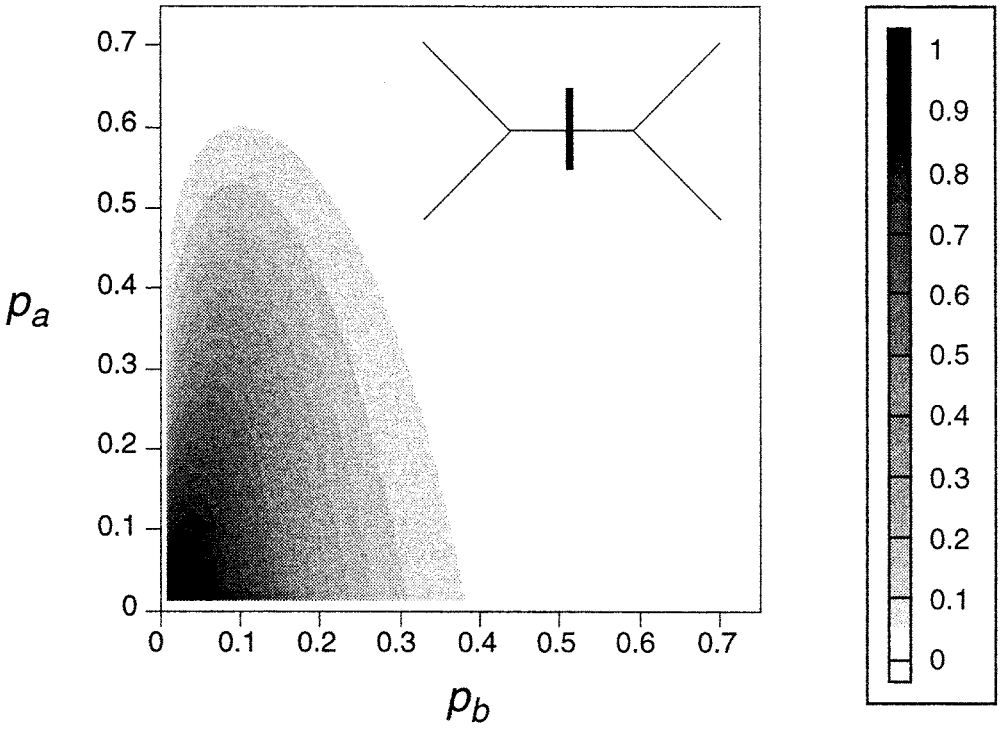

FIGURE 6. Contour plots showing the proportion of parsimony-informative sites that represent true synapomorphies (a single change along the internal branch) with no changes in the terminal branch. In the extreme regions of the inverse-Felsenstein zone (upper left corner), nearly all of the parsimony-informative characters support the true tree, but almost none of them will be true synapomorphies. See Figure 1 for definition of p_a and p_b.

in Figure 5b. Figure 6 illustrates the bottom line; almost all of the good performance by parsimony in the inverse-Felsenstein zone is due to sites with more than one substitution. Siddall (citing Farris, 1983) was apparently aware that parsimony's performance was being boosted by misinterpreted homoplasy, as suggested by the following statement (Siddall, 1998:216): "the reason [that parsimony does well in the inverse-Felsenstein zone] 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." We would not dispute this statement in the least. However, we would add that most researchers would be worried if they knew that 99% (or more) of the apparent support for an "optimal" tree came from an inherent bias in the method used rather than from actual phylogenetic signal. Surprisingly, Siddall seems entirely comfortable with this possibility, referring to parsimony as "positively leading" (1998:216) in the inverse-Felsenstein zone.

The ultimate cause of the bias toward trees that group "long-branch" taxa is that parsimony severely underestimates the true number of substitutions that occur along the long branches. It is important to remember that likelihood methods can have similar problems when their models are strongly violated. In general, if the violation of the model is such that the assumed model is too simple (e.g., if high transition/transversion ratios or among-site rate variations are ignored), underestimation of the actual number of substitutions can lead to inconsistency of likelihood in the Felsenstein zone (Waddell, 1995:377–385; Gaut and Lewis, 1995; Chang, 1996a; Sullivan and Swofford, 1997) and overconfidence in the inverse-Felsenstein zone (Waddell, 1995:385–398; Bruno and Halpern, 1999). However, attempting to account for multiple substitutions by using an oversimplified model is a step in the right direction, whereas ignoring them entirely is to accept ignorance. Maximum likelihood methods are much more robust to artifacts

of long-branch attraction than are parsimony methods, even when their assumed models are inadequate.

Long-Branch Repulsion or Absence of Long-Branch Attraction?

If the true unrooted tree for four taxa has an internal branch length close enough to zero that the information in the resulting sequences is insufficient to reliably choose one of the trees over the others, then the tree is effectively a star tree. In this case, if a method is unbiased, it should choose equivocally—it might favor all three trees equally, or it might choose one tree at random (and, ideally, discover as well that the other trees were not significantly different). By this argument, if a method correctly chooses the true tree one-third of the time, then it is successful, even though it chooses an incorrect topology the other two-thirds of the time. On the other hand, if a biased method is used when the true tree is effectively a star tree, one topology will be preferred over the others. If there are exactly three choices and the available information is inadequate to decide among them, then the method is failing if it deviates strongly from a 1 in 3 preference for each choice. In this case, a method obviously is failing if it preferentially chooses the wrong tree, but perhaps less obviously, it is also failing if it always favors the correct tree.

Siddall (1998) focused on a near-star tree in the inverse-Felsenstein zone for which there was little or no information in the sequences to distinguish among the three possible trees and found that parsimony nonetheless chooses the correct tree topology most of the time. A similar situation exists for a near-star tree in the Felsenstein zone, except that parsimony usually chooses the same topology, which is now incorrect. In both cases, parsimony is failing rather than succeeding, its failure following directly from its bias. Likelihood, on the other hand, is succeeding in both of these cases because its choice is much closer to a random one in both zones. Virtually any method for assessing reliability, including bootstrapping (Felsenstein, 1985), jackknifing (Penny and Hendy, 1986; Felsenstein, 1988; Farris et al., 1996), or the test of Kishino and Hasegawa (1989), will fail to find significant support for any of the three possible topologies under the likelihood criterion. In this case, it is not appropriate to say the likelihood is failing because of "long-branch repulsion;" rather, it is succeeding in remaining uncommitted when the data do not decisively support any single topology. If indeed likelihood were affected by long-branch repulsion, then it would obtain the correct topology significantly less than one-third of the time, which it does not do.

This basic notion can be encapsulated in the simulation results shown in Figure 7. This simulation evaluates the relative performance of parsimony and likelihood for three sequence lengths as the tree approaches a star tree from the inverse-Felsenstein zone, becomes an exact star tree, and then moves into the Felsenstein zone. Likelihood methods do well in both zones when the central branch length is at least 0.04 substitutions per site and the number of sites is not small. Parsimony is inconsistent in the Felsenstein zone for all branch lengths in the range of 0–0.05 substitutions per site (and even higher), doing better for shorter sequences than longer ones. As the central branch length shrinks toward zero in both zones, the accuracy of likelihood decreases, reaching the expected 1 in 3 accuracy rate when the central branch is extremely small in either zone. Bootstrap (as well as jackknife) support is low for all three trees (details not shown). Parsimony, on the other hand, abruptly shifts from nearly perfect accuracy to complete inaccuracy on either side of the zero point. When the central branch is extremely short, parsimony simply chooses the tree that groups the long-branch taxa, regardless of what the true tree might be, with high bootstrap support for either the correct or the incorrect result. This behavior of parsimony in the extreme regions of the Felsenstein and inverse-Felsenstein zones is analogous to an oracle who responds to any question by responding "0.492." If the question asked is, "What is the sum of 0.450 and 0.042?" or "What is 3 times 0.164?" the oracle will answer correctly, but presumably once interrogators realized that the answer was always the same regardless of the question, they would not be ready to give up their electronic calculators. There are times when "I don't know" is a better answer than a confident guess that has a high probability of being incorrect.

It is interesting, and somewhat amusing, to examine the performance of a phenetic

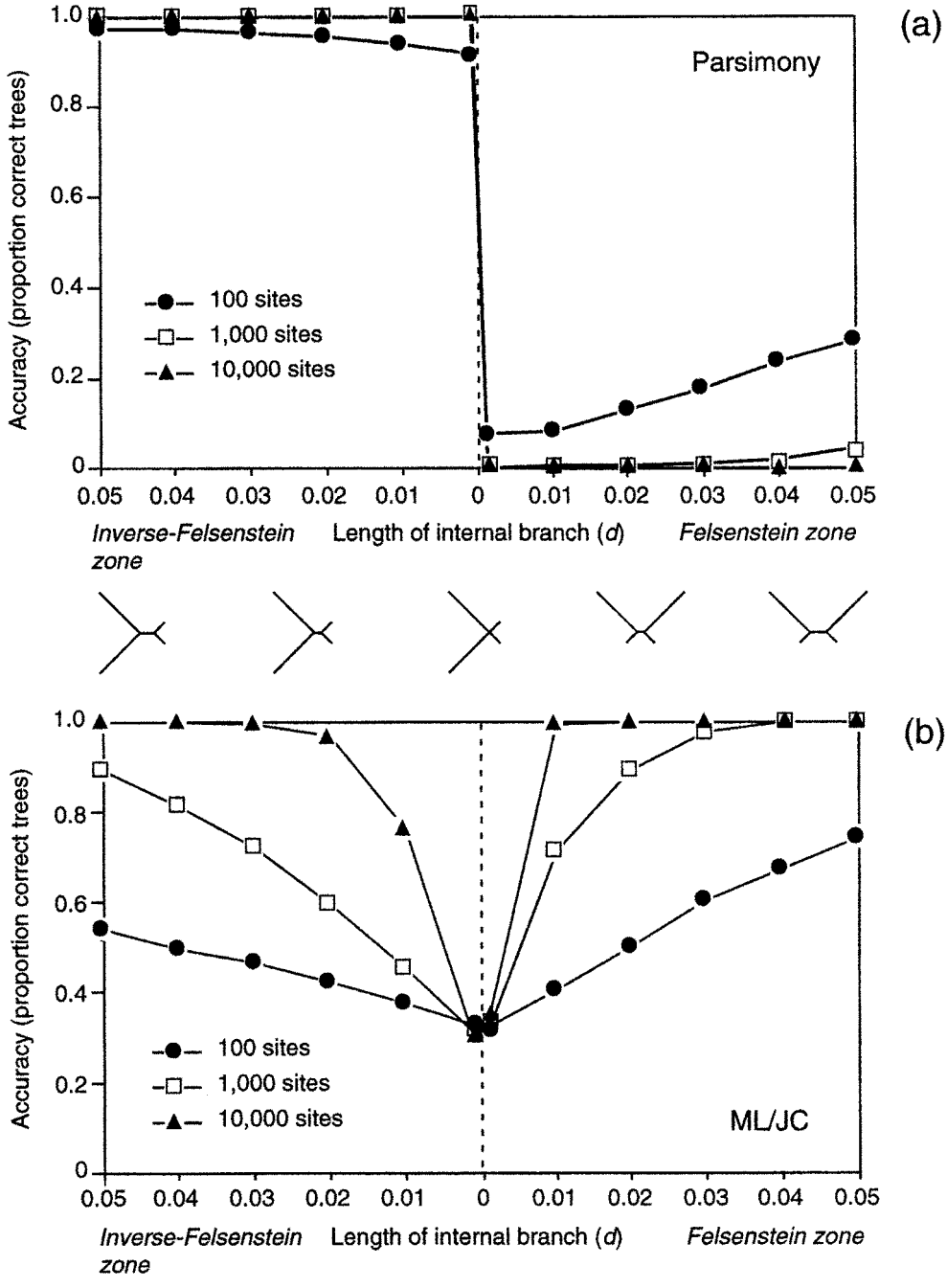

FIGURE 7. Results of simulation comparing the behavior of parsimony with likelihood in the transition between the inverse-Felsenstein and Felsenstein zones. The lengths of the terminal branches, in expected substitutions per site, are 0.5 (long branches) and 0.05 (short branches). Accuracy is measured as the proportion of correctly estimated trees in 1,000 simulation replicates. (a) Parsimony shifts abruptly from nearly perfect accuracy to nearly complete inaccuracy as the true tree goes from being a near-star tree in the inverse-Felsenstein zone to a near-star tree in the Felsenstein zone, especially for sequence lengths of 1,000 or longer. (b) When the internal branch length is not extremely small or the sequence length is not too short, likelihood achieves reasonable accuracy in both the inverse-Felsenstein and Felsenstein zones. As the internal branch becomes progressively shorter, likelihood is less able to infer the tree correctly, appropriately reflecting the lack of resolving power in the data. In both plots, the points just to the left and right of the zero on the abscissa represent branches of infinitesimal length in the inverse-Felsenstein and Felsenstein zones, respectively.

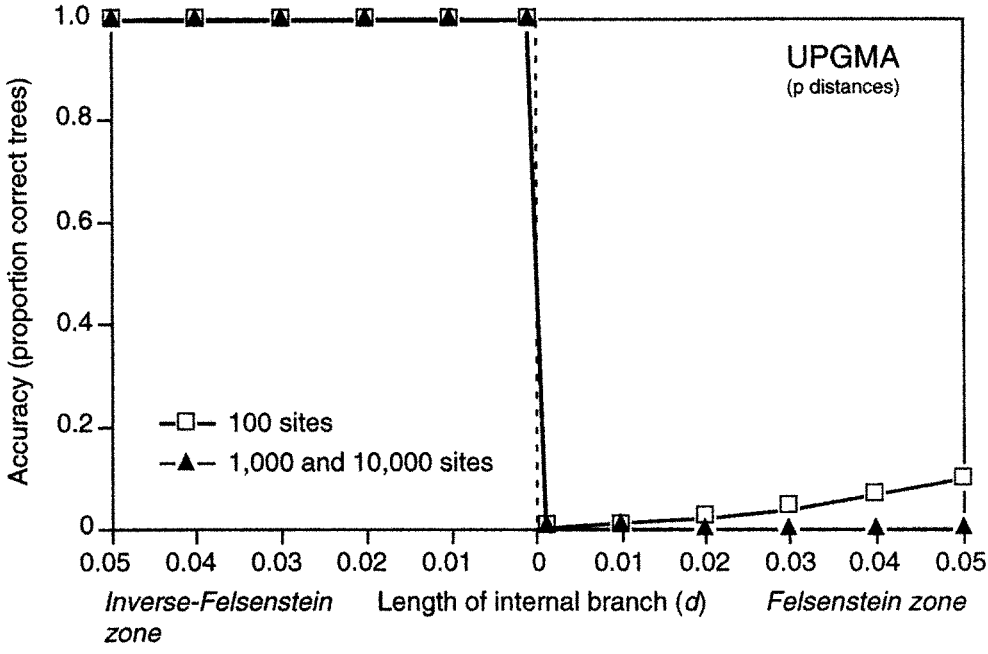

FIGURE 8. Performance of UPGMA in the transition between the inverse-Felsenstein and Felsenstein zones mimics that of parsimony (see Fig. 7a), reinforcing the position that the high "accuracy" of parsimony in the inverse-Felsenstein zone is purely the result of bias.

clustering method, UPGMA, under the same conditions (Fig. 8). Because UPGMA determines a rooted tree, we can disregard the position of the root in order to compare its performance fairly with that of the intrinsically unrooted parsimony and maximum likelihood methods. UPGMA demonstrates almost exactly the same behavior as parsimony. It finds the correct tree 100% of the time in the inverse-Felsenstein zone, at the price of missing it nearly 100% of the time in the Felsenstein zone. Application of UPGMA as a phylogenetic method requires the assumption of a "molecular clock" (equal rates of substitution in all lineages), but violation of this assumption does not necessarily lead to reduced accuracy—the method's inherent bias works in its favor if the taxa that are most similar are in fact close relatives. Reasoning analogous to Siddall's could then be used to argue that UPGMA is behaving properly (even outperforming parsimony) by avoiding "long-branch repulsion" in the inverse-Felsenstein zone when the clock assumption is violated. We doubt, however, that many proponents of parsimony methods would find this argument compelling.

CONCLUSIONS

The demonstration that parsimony analysis can, under specific conditions, achieve greater accuracy than maximum likelihood fails to rescue parsimony from the criticism that potential biases can lead it to support or reject alternative topologies strongly when information is insufficient for reaching a definitive conclusion. Although the property of providing strong support for a *correct*, but near-star, tree in the inverse-Felsenstein zone seems desirable, this advantage is negated if the method also provides strong support for an *incorrect*, but near-star, tree in the Felsenstein zone. Whatever good properties parsimony might have—and we do not deny their existence—strong commitment to a topology purely on the basis of the length of a tree's terminal branches is not one of them.

It is often suggested that the conditions under which maximum likelihood outperforms parsimony are extreme, whereas under more "typical" conditions this advantage disappears. Siddall has taken a different position, claiming to have found a "limiting case" for which likelihood methods, rather than parsimony, "more often than not will fail to converge on the correct model topology"

(Siddall, 1998:211). We have shown that this claim is simply false and further suggest that most scientists would prefer to use methods that are honest about how strongly a result is supported than to use a method that pretends that a result is strongly supported when the majority of that support is a consequence of bias. When interpreted properly, simulation studies such as those of Siddall (1998) merely reinforce arguments for the utility of model-based methods, including maximum likelihood, for phylogenetic analysis of molecular sequence data. These methods acknowledge the inevitability of multiple substitutions and explicitly accommodate them as a fundamental component of their operation. The parsimony method is useful and powerful in many situations, but its ability to obtain a "correct" result for reasons that are clearly inappropriate should not be used as an argument in its favor.

ACKNOWLEDGMENTS

We thank Hirohisa Kishino, Jack Sullivan, Jianzhi Zhang, and members of the "Phylobrew" group (especially Frank "Andy" Anderson, Robb Brumfield, Kevin de Queiroz, Jim McGuire, Steve Poe, and Jim Wilgenbusch) for helpful discussion and editorial suggestions. Correspondence with Ziheng Yang was extremely helpful in clarifying our ideas on Wald's consistency proof and related ideas, although this should not be taken as complete endorsement of what we have written here. We thank Ron DeBry, Mark Siddall, and an anonymous reviewer for suggestions that improved the clarity and accuracy of the paper. Finally, the "oracle" analogy is not original to this paper. We have heard variations on it from Joe Felsenstein and Bret Larget, among others.

This work was supported by the following grants: Marsden Fund of New Zealand to P.J.W., National Science Foundation (NSF) DEB-0075406 to J.P.H., NSF DEB-9628835 to J.S.R., NSF DEB-9974124 to D.L.S., and an Alfred P. Sloan/NSF Young Investigator Award to P.O.L.

REFERENCES

BRUNO, W. J., AND A. L. HALPERN. 1999. Topological bias and inconsistency of maximum likelihood using wrong models. Mol. Biol. Evol. 16:564–566.

CHANG, J. T. 1996a. Inconsistency of evolutionary tree topology reconstruction methods when substitution rates vary across characters. Math. Biosci. 134:189–215.

CHANG, J. T. 1996b. Full reconstruction of Markov models on evolutionary trees: Identifiability and consistency. Math. Biosci. 137:51–73.

FARRIS, J. S. 1983. The logical basis of phylogenetic analysis. Pages 7–36 *in* Advances in cladistics, Volume 2 (N. I. Platnick and V. A. Funk, eds.). Columbia Univ. Press, New York.

FARRIS, J. S. 1997. "Who, really is a statistician?" Paper presented at the Sixteenth Meeting of the Willi Hennig Society. George Washington Univ. Washington, DC.

FARRIS, J. S. 1999. Likelihood and inconsistency. Cladistics 15:199–204.

FARRIS, J. S., V. A. ALBERT, D. LIPSCOMB, AND A. G. Kluge. 1996. Parsimony jackknifing outperforms neighbor-joining. Cladistics 12:99–124.

FELSENSTEIN, J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. Syst. Zool. 27:401–410.

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 P. O. LEWIS. 1995. Success of maximum likelihood in the four-taxon case. Mol. Biol. Evol. 12:152–162.

HENDY, M. D., and D. PENNY. 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. 38:297–309.

HILLIS, D. M., HUELSENBECK, J. P. AND D. L. SWOFFORD. 1994. Hobgoblin of phylogenetics? Nature 369:363–364.

HUELSENBECK, J. 1995. Performance of phylogenetic methods in simulation. Syst. Biol. 44:17–48.

HUELSENBECK, J. P., AND D. M. HILLIS. 1993. Success of phylogenetic methods in the four-taxon case. Syst. Zool. 42:247–264.

JUKES, T. H., AND C. R. CANTOR. 1969. Evolution of protein molecules. Pages 21–132 *in* Mammalian protein metabolism. (H. N. Munro, ed.). Academic Press, New York.

KENDALL, M., AND A. STUART. 1979. Advanced theory of statistics, 2nd edition. Charles Griffin, London.

KIM, J. 1996. General inconsistency conditions for maximum parsimony: Effects of branch lengths and increasing numbers of taxa. Syst. Biol. 45:363–374.

KISHINO, H., AND M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. 29:170–179.

KUHNER, M. K., and J. FELSENSTEIN. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Mol. Biol. Evol. 11:459–468.

LOCKHART, P. J., A. W. LARKUM, M. A. STEEL, P. J. WADDELL, AND D. PENNY. 1996. Evolution of chlorophyll and bacteriochlorophyll: The problem of invariant sites in sequence analysis. Proc. Natl. Acad. Sci. USA 93:1930–1934.

PENNY, D., AND M. D. HENDY. 1986. Estimating the reliability of evolutionary trees. Mol. Biol. Evol. 3:403–417.

ROGERS, J. S. 1997. On the consistency of maximum likelihood estimation of phylogenetic trees from nucleotide sequences. Syst. Biol. 46:354–357.

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 A. G. KLUGE. 1997. Probabilism and phylogenetic inference. Cladistics 13:313–336.

SIDOW, A. 1994. Parsimony or statistics? Nature 367:26–27.

SULLIVAN, J., AND D. L. SWOFFORD. 1997. Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. J. Mammal. Evol. 4:77–86.

SWOFFORD, D. L., P. O. LEWIS, AND P. J. WADDELL. 1995. "The phylogenetic utility of LogDet/paralinear distances for more realistic evolutionary models. I. Is there a heavy price for using them when a simpler model would suffice?" Paper presented at annual meeting of the Society for the Study of Evolution/Society of Systematic Biologists, Montreal, Canada.

WADDELL, P. J. 1995. Statistical methods of phylogenetic analysis, including Hadamard conjugations, LogDet transforms, and maximum likelihood. Massey Univ., Palmerston North, New Zealand.

WADDELL, P. J., D. PENNY, AND T. MOORE. 1997. Hadamard conjugations and modeling sequence evolution with unequal rates across sites. Mol. Phylogenet. Evol. 8:33–50.

WALD, A. 1949. Note on the consistency of maximum likelihood. Ann. Math. Stat. 20:595–601.

YANG, Z. 1996. Phylogenetic analysis using parsimony and likelihood methods. J. Mol. Evol. 42:294–307.

Received 19 January 2000; accepted 18 March 2000
Associate Editor: R. Olmstead