The Consistency of Several Phylogeny-Inference Methods under Varying Evolutionary Rates[1]

Ronald W. DeBry
Department of Chemistry, Florida State University

A phylogenetic method is a consistent estimator of phylogeny if and only if it is guaranteed to give the correct tree, given that sufficient (possibly infinite) independent data are examined. The following methods are examined for consistency: UPGMA (unweighted pair-group method, averages), NJ (neighbor joining), MF (modified Farris), and P (parsimony). A two-parameter model of nucleotide sequence substitution is used, and the expected distribution of character states is calculated. Without perfect correction for superimposed substitutions, all four methods may be inconsistent if there is but one branch evolving at a faster rate than the other branches. Partial correction of observed distances improves the robustness of the NJ method to rate variation, and perfect correction makes the NJ method a consistent estimator for all combinations of rates that were examined. The sensitivity of all the methods to unequal rates varies over a wide range, so relative-rate tests are unlikely to be a reliable guide for accepting or rejecting phylogenies based on parsimony analysis.

Downloaded from https://academic.oup.com/mbe/article/3/3/173/1007296 by guest on 16 April 2021

Introduction

Phylogenetic analysis should be viewed as a problem of statistical inference (Felsenstein 1983; Rohlf and Wooten 1988; Goldman 1990), but the methods most commonly used to estimate phylogeny have at best an uncertain foundation in statistical theory. This does not mean that these methods should not be used, but more attention must be paid to their behavior as statistical estimators under different models of evolution. To some extent, some methods, such as parsimony (Edwards and Cavalli-Sforza 1964; Eck and Dayhoff 1966, p. 164; Fitch 1971), have been studied in a statistical framework. Felsenstein (1973) first showed that parsimony is not the same as maximum-likelihood estimation (which is an explicitly statistical method); later, Felsenstein (1978) and Cavender (1978) demonstrated that parsimony, when based on discrete characters, can fail, under certain conditions, to be a consistent estimator of the phylogeny. A statistic has the property of consistency if it converges on the correct value as more and more data are collected. These findings mean that workers using discrete character parsimony must be willing to accept certain assumptions about the nature of the evolutionary process, since, under at least some conditions, parsimony will converge on the wrong phylogeny. Potentially the most important assumption, according to previous work (Cavender 1978; Felsenstein 1978; Hendy and Penny 1989), is that rates of evolution are either nearly equal among taxa or very low overall. It is not known how often parsimony might be inconsistent in the real

1. Key words: consistency, maximum-parsimony method, modified Farris method, neighbor-joining method.

Address for correspondence and reprints: Ronald W. DeBry, Department of Chemistry, Florida State University, Tallahassee, Florida 32306.

Mol. Biol. Evol. 9(3):537–551. 1992.
© 1992 by The University of Chicago. All rights reserved.
0737-4038/92/0903-0013$02.00

world, and it is not known whether other commonly used methods are consistent estimators when parsimony is not. To further complicate matters, Hendy and Penny (1989) have recently shown a particular case where parsimony can fail to be consistent even if the rates of evolution are exactly clocklike. Distance methods seeking a minimum-length tree have also been developed for distance data. These methods should be consistent if rates vary among taxa, provided that the distances are accurately corrected for multiple substitutions (Felsenstein 1988). Perfect distance correction is probably an unattainable goal, however, and it may be that, under the same conditions that threaten discrete character parsimony, imperfectly corrected distances can result in inconsistency.

In the present paper I study three parsimony methods: the discrete-character parsimony (P) method and two distance-parsimony methods, a slight modification of Farris' (1972) distance Wagner method, known as the modified Farris (MF) method (Tateno et al. 1982), and the neighbor-joining (NJ) method of Saitou and Nei (1987). Both the MF and NJ algorithms build the tree according to a rigid stepwise procedure, minimizing the total length of the tree at each step. The methods differ in that the NJ method minimizes the length of the tree over all taxa at each step, while the MF method only minimizes the total length of the subtree containing the taxon added at that stage. Neither method attempts to adjust the tree to minimize some measure of difference between distances calculated from the tree and those observed. In the present paper I focus on the evolution of nucleotide sequence characters and examine the effect of using either no correction or the Jukes-Cantor (Jukes and Cantor 1969) one-parameter correction on data that are generated using the two-parameter model of Kimura (1980). The two-parameter model is in many cases biologically realistic for DNA sequence data (at least as a first approximation), because it describes the situation where the rate of transition substitutions is different from the rate of transversion substitutions. In the present study the transition:transversion ratio is always assumed to be 9:1. In addition to the P, MF, and NJ methods, a UPGMA (unweighted pair-group method, averages) (Sneath and Sokal 1973, p. 230) phenogram is constructed for each scenario, on the basis of the uncorrected distances. UPGMA is widely described as a method that requires equal rates among taxa, in order to produce the correct tree (e.g., see Colless 1970), and so provides a baseline for examining how robust the other methods are to violations of an equal rate assumption.

For each scenario, I calculate both the expected final distribution of character states and the expected distance matrix, given the model and branch lengths. This is equivalent to having an infinitely long DNA sequence that evolves exactly as predicted by the probability model, so there are no effects due to sampling errors. This is an artificial situation, but it is a useful approach to take when one asks the question, Under what conditions will a method be a consistent estimator?

Methods

Only five taxa are investigated in the present study. Most systematic studies include at least one taxon which is known or presumed to be an outgroup, so this will be mimicked here. When an explicit outgroup is used to root the trees, there are 15 possible topologies for each of the four methods, so the different methods can be compared directly.

It is possible to determine consistency under these simple models by simulation studies, but for small numbers of taxa it is practical to calculate the expected frequencies of all possible data outcomes, as done by Felsenstein (1978). Felsenstein's approach

Downloaded from https://academic.oup.com/mbe/article/8/3/537/1007266 by guest on 19 April 2024

is to calculate the probability of each possible character-state outcome, given an assumed topology and set of branch lengths (an "evolutionary scenario"). If the model were exactly correct, or if the scenario were used to generate characters in a simulation, then the observed frequencies of the character states would converge on these probabilities as an infinitely large data set was accumulated. A method will be consistent for a particular scenario if it gives the correct topology when applied to a hypothetical data set that exactly matches the expected character distribution. I begin with a five-taxon topology and assume the operation of a perfect molecular clock, and I then increase the rate of nucleotide substitution in just one of the terminal branches, until each method becomes inconsistent.

Calculation of Substitution Probabilities

Kimura's (1980) two-parameter model of nucleotide sequence evolution was used. Under this model, different rates of transitions and transversions are allowed. The total rate of change per nucleotide site per unit time is $\lambda = \alpha + 2\beta$, where α is the rate of transitions and β is the rate of each of the two kinds of transversions. In the present study, the ratio between the rates of transitions and transversions is fixed, so that $\alpha:2\beta = 9:1$. The continuous-time solutions to this model are provided by Li [1986, eq. (1)], who follows Aoki et al. (1981). If $P_A(0)$, $P_T(0)$, $P_C(0)$, and $P_G(0)$ are the probabilities that the base at a particular position at the beginning of a branch with length t is A, T, C, and G, respectively, then the probabilities that the base will be A, T, C, or G at time t are, respectively,

$$P_A(t) = \tfrac{1}{4} + \tfrac{1}{2}[P_A(0) + P_G(0) - \tfrac{1}{2}]e^{-4\beta t} + \tfrac{1}{2}[P_A(0) - P_G(0)]e^{-2(\alpha+\beta)t} \quad ; \quad (1a)$$

$$P_T(t) = \tfrac{1}{4} + \tfrac{1}{2}[P_T(0) + P_C(0) - \tfrac{1}{2}]e^{-4\beta t} + \tfrac{1}{2}[P_T(0) - P_C(0)]e^{-2(\alpha+\beta)t} \quad ; \quad (1b)$$

$$P_C(t) = \tfrac{1}{4} + \tfrac{1}{2}[P_C(0) + P_T(0) - \tfrac{1}{2}]e^{-4\beta t} + \tfrac{1}{2}[P_C(0) - P_T(0)]e^{-2(\alpha+\beta)t} \quad ; \quad (1c)$$

$$P_G(t) = \tfrac{1}{4} + \tfrac{1}{2}[P_G(0) + P_A(0) - \tfrac{1}{2}]e^{-4\beta t} + \tfrac{1}{2}[P_G(0) - P_A(0)]e^{-2(\alpha+\beta)t} \quad . \quad (1d)$$

Calculation of Data Outcome Probabilities

For a system with five taxa and four character states, there are $4^5 = 1{,}024$ possible data outcomes at any particular nucleotide site. Given an evolutionary scenario and Li's (1986) model, we can calculate the probability of observing the ith data outcome as

$$P_i = \sum_{j=1} \prod_{k=1} P_{s_k s_{k'}}(t_k) \,, \quad (2)$$

where the summation (j) is over all possible assignments of states to each of the interior nodes, and $P_{s_k s_{k'}}(t_k)$ is the probability of a character changing from starting state s_k to ending state $s_{k'}$ over a branch with length t_k (Felsenstein 1979). For interior branches, $s_{k'}$ takes on all four possible values, but, for each terminal branch, $s_{k'}$ is fixed to the value determined by i.

Calculation of Expected Tree Topologies

If this model were used in a simulation study, then, as larger and larger data sets were generated, the observed frequency of each data type would converge on the

Downloaded from https://academic.oup.com/mbe/article/3/6/537/1027266 by guest on 20 April 2021

expected values obtained from equation (2) (Felsenstein 1978). Therefore, the tree topology inferred by each method from the expected character distributions represents the topology to which that method would converge. The topology used to generate the expected character distributions for five species is shown in figure 1. If the expected preferred topology for a scenario is the same as the true topology, then that method is a consistent estimator of phylogeny for that set of parameters. The expected character-state frequencies are easily converted into an expected distance matrix. The entries in this matrix represent the proportion of nucleotide positions that differ between two species. Details of the algorithms used to calculate the tree topologies for each method were obtained from the literature.

Having determined the probabilities of each of the 1,024 different character distributions, we can find the expected number of steps for each of the 15 topologies. The topology with the lowest expected number of steps is the expected topology of the P method. First, a matrix S is constructed with 15 columns, one for each topology, and with 1,024 rows, one for each different character-state outcome. Each S_{ij} is the minimum number of steps required to map data outcome i onto topology j. The probability of obtaining the ith-data outcome is also the expected proportion of char-

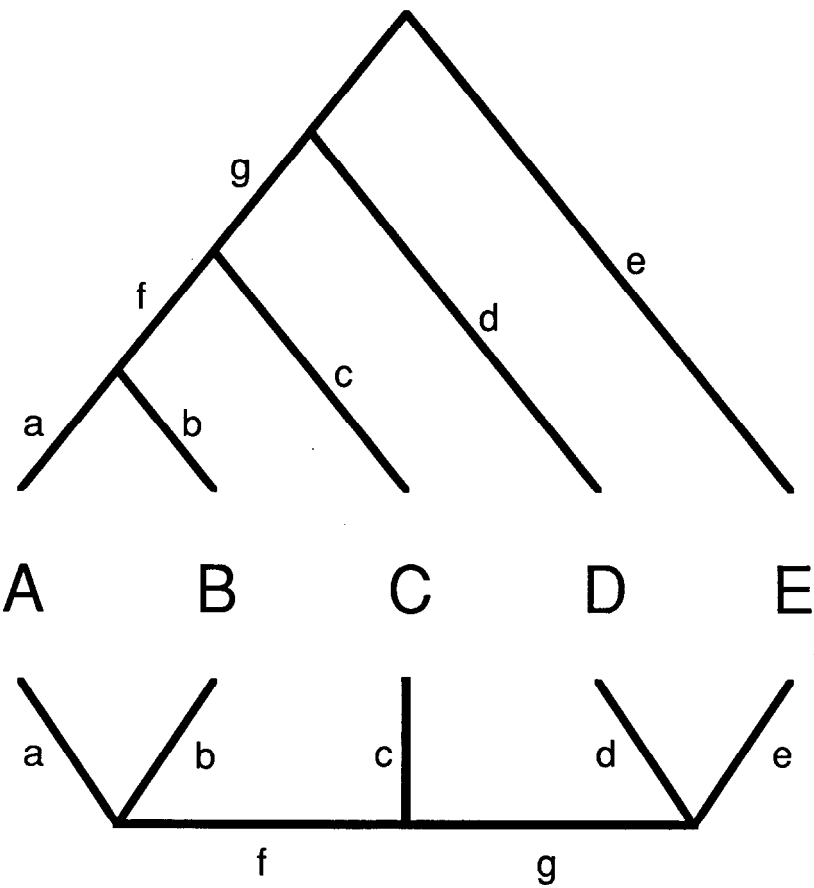


FIG. 1.—Relationships among sequences used in present study. Uppercase letters represent terminal taxa, and lowercase letters represent branches and their corresponding lengths. One possible rooting of the unrooted (lower) tree is shown in the upper tree. This rooting is used for most of the cases studied.

Downloaded from https://academic.oup.com/mbe/article/9/3/537/1037266 by guest on 10 April 2024

acters with that outcome. If **p** is the row vector containing these 1,024 probabilities, then a weighted-steps matrix, **W**, may be obtained as **W** = **pS**. The expected length (per character) of the *j*th topology is given by summing the expected lengths contributed by each data-outcome pattern:

$$\text{Exp}(L_j) = \sum_{i=1} W_{ij} . \quad (3)$$

If the true topology (shown in fig. 1) is assigned number 1 (of 15), then parsimony is a consistent estimator for that particular scenario if $\text{Exp}(L_1)$ is smaller than all the remaining $\text{Exp}(L_j)$.

Results

Figure 1 shows the tree topology used for the following calculations; the unrooted topology shown in the lower part of figure 1 is used for all the scenarios. To impose a molecular clock, it is necessary to specify a root. Rooting along branch e gives the tree shown in the upper part of figure 1. This root results in an asymmetrical branching pattern within the ingroup and is used for most of the scenarios. In one case, however, the tree is rooted on branch c, giving a symmetrical branching pattern within the ingroup. For each particular case examined, the branch lengths and rooting are given in table 1.

Felsenstein (1978) showed that, under his simple model, the robustness of parsimony to rate variation among taxa decreased as the overall rate of evolution increased. Figure 2 shows the results obtained as the total length of the tree is increased while the proportions of the tree remain the same. In figures 2–4, the topological parameter that is being varied is shown on the *X*-axis. For each value of that parameter, only one branch at a time is allowed to have a substitution rate that departs from the molecular clock. In figure 2, the nodes are equidistant, with the lengths of internal branches f and g set equal to the length of the terminal branch a. For each value of a, the length of one terminal branch, b (fig. 2A) or c (fig. 2B), is increased until each of the methods becomes inconsistent. Those lengths are expressed as a multiple of the original, clocklike branch length and are plotted along the *Y*-axis. For example, in

Table 1
Branch Lengths for Cases Studied

Case	Length at Branch[a]							Outgroup	Figure
	a	b	c	d	e	f	g		
1	a	b	2a	3a	10a	a	a	E	2a
2	a	a	c	3a	10a	a	a	E	2b
3	0.05	b	0.10	0.15	e	0.05	0.05	E	3a
4	0.05	0.05	c	0.15	e	0.05	0.05	E	3b
5	a	b	a + f	2a + f	10a	f	a	E	4
6	a	a	3(a + f)	a	a	f	f	C	5

[a] a–e represent lengths of the terminal branches leading to sequences A–E, respectively, which are related as shown in fig. 1. f and g represent the lengths of the interior branches in fig. 1. For the various cases examined, lengths are given for branches that are held constant. For branches that vary within a case, lengths are given in terms of branches a and f. The branch used to root the tree and the figure showing the plots for each case are provided.

Downloaded from https://academic.oup.com/mbe/article/9/5/537/1038266 by guest on 10 April 2024

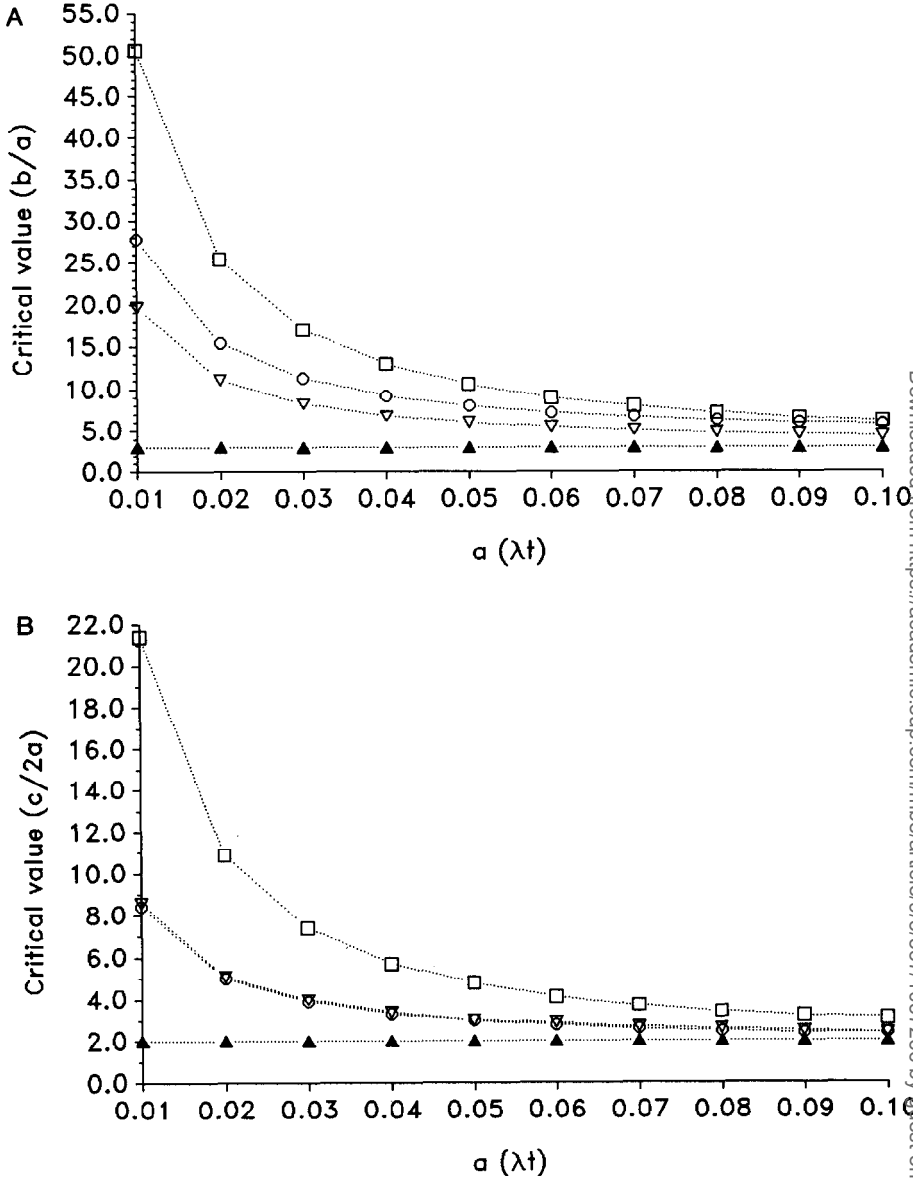


FIG. 2.—Relationship between total tree length and consistency, when only one branch is allowed to deviate from molecular clock. The tree shown in fig. 1 is rooted along branch e, with branch lengths as given in table 1. Squares denote values for the NJ method using Jukes-Cantor corrected distances; circles denote values from P; unblackened triangles denote values from NJ using uncorrected distances; and blackened triangles denote values from UPGMA. As shown in panel A, for a particular method and value of a, the critical value is the length of branch b (expressed as a multiple of its length under the molecular clock) at which the correct tree is no longer chosen by that method. Thus, below its critical value each method is a consistent estimator of the correct tree, and above the critical value it is not. When branch b is lengthened, critical values for the MF method are identical to those for UPGMA. Panel B is the same as panel A, except that c is the only branch that is not constrained by the molecular clock. When branch c is lengthened, critical values for the MF method are identical to those for the NJ method.

Downloaded from https://academic.oup.com/mbe/article/9/3/537/1037266 by guest on 16 August 2022

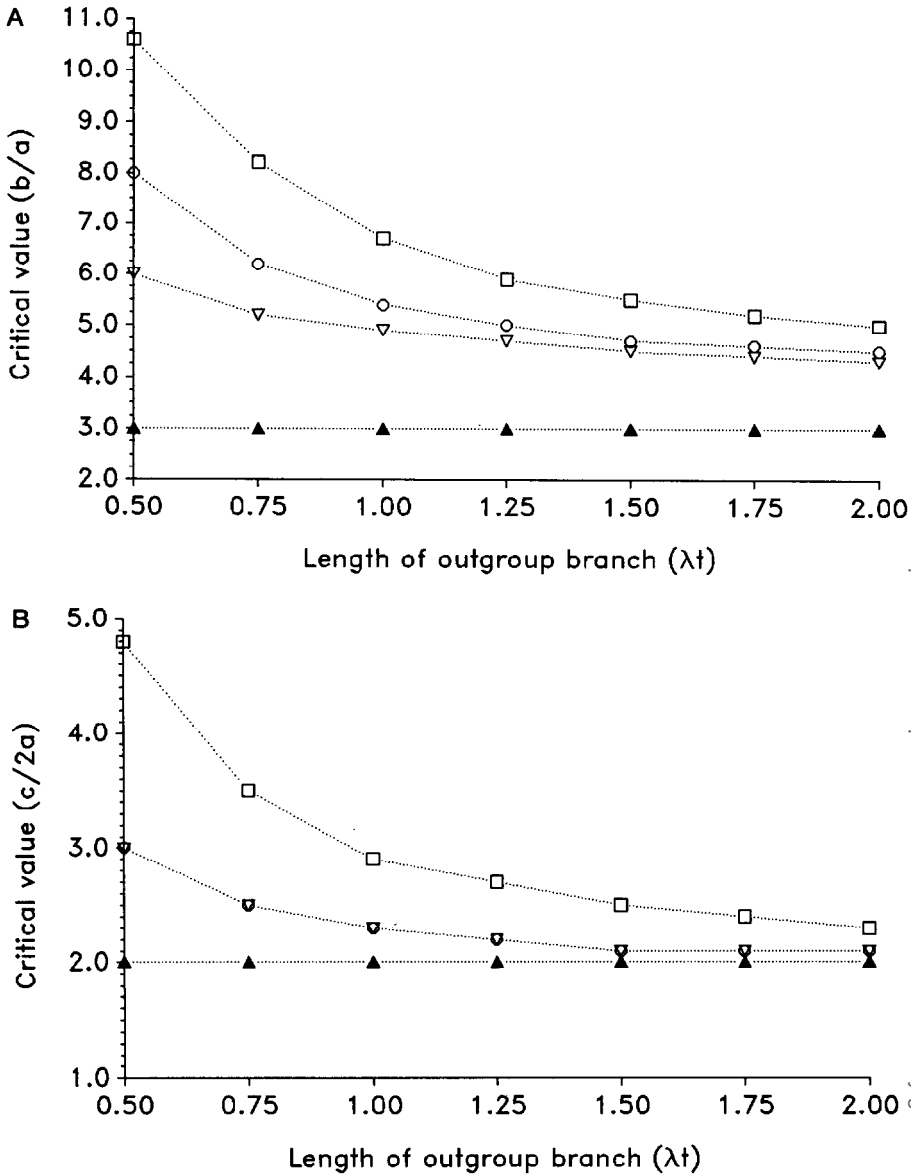
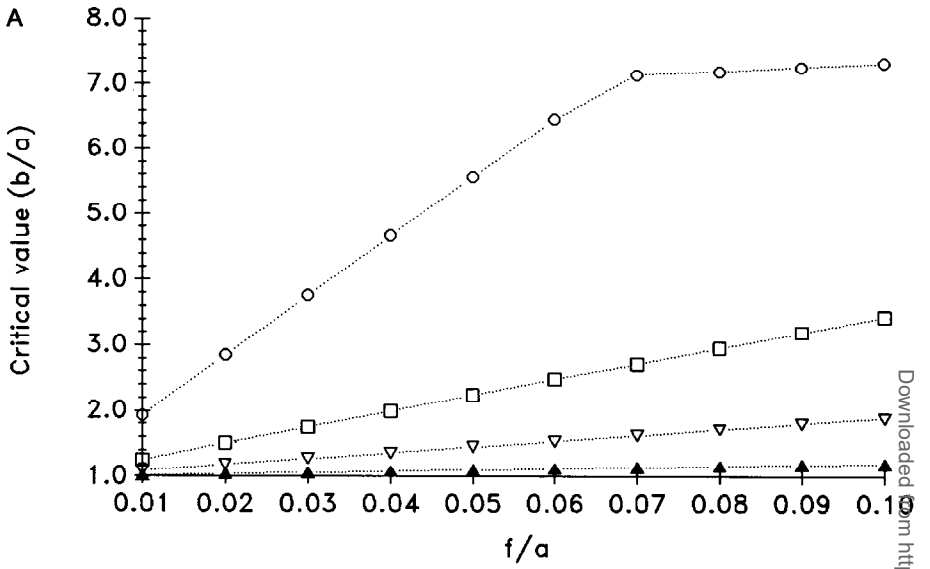


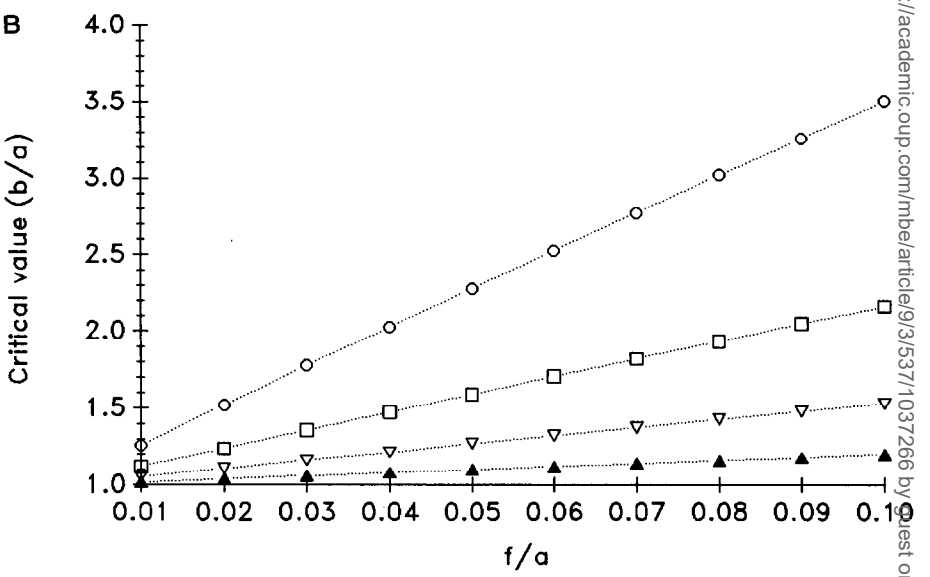
FIG. 3.—Relationship between length of branch leading to outgroup and consistency, when only one branch is allowed to deviate from molecular clock. Symbols are the same as in fig. 2, and branch lengths are as given in table 1. A, Critical values when length of branch b is increased relative to its length under molecular clock. Critical values for the MF method are identical to those for UPGMA. B, Critical values when length of branch c is increased relative to its length under molecular clock. Critical values for the MF method are identical to those for the NJ method.

figure 2A, when a = 0.01, b must exceed 0.28 for the P method to be inconsistent, a 28-fold increase in substitution rate. On the same starting tree, c (which starts with a length of 0.02 when a = 0.01) must be lengthened by just over eightfold to cause the P method to choose an incorrect topology.

When the one-parameter model is used to generate the expected character-dis-

Downloaded from https://academic.oup.com/mbe/article/9/3/537/1037266 by guest on 16 August 2022





Downloaded from https://academic.oup.com/mbe/article/9/3/537/1037266 by guest on 10 April 2024

tribution spectrum, and when the Jukes-Cantor correction (which gives the true distances under the one-parameter model) is applied to the distance matrix, then the NJ method is consistent for every set of branch lengths tried (results not shown). But, when the observed distances are not properly corrected for multiple substitutions, the relationship that the NJ method shows between rate variation and consistency is similar to that for the P method. The MF method shows an unexpected behavior. When c is varied, the MF method always chooses the same topology as does the NJ method, regardless of which model is used to generate the expectations or which correction is applied to the expected distances. But, when b is lengthened, the MF method always gives the same topology as UPGMA, again regardless of which model or correction is used. When one looks at the intermediate results as taxa are added to the MF tree

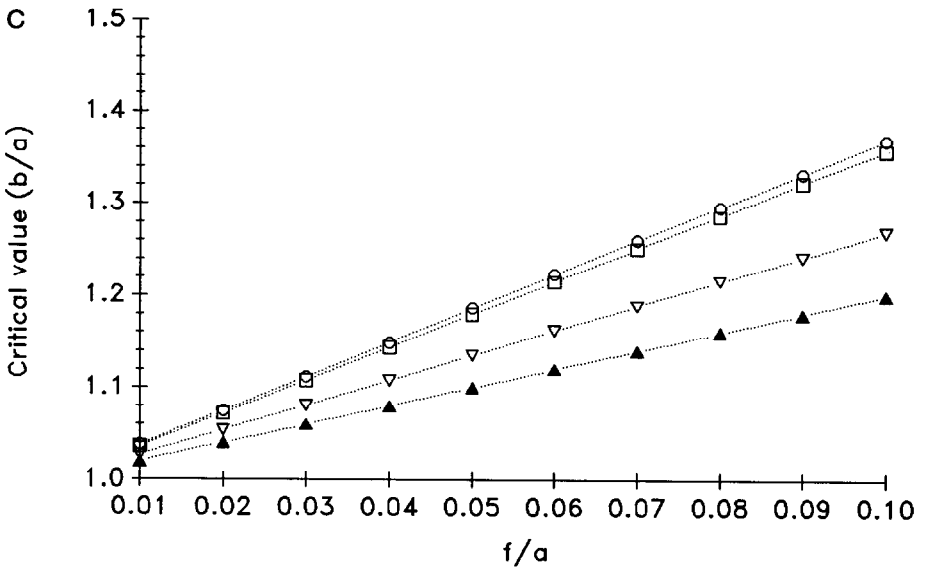


FIG. 4.—Relationship between length of internal branch f and consistency, when only branch b is allowed to deviate from molecular clock. Symbols are the same as in fig. 2, and branch lengths are as given in table 1. A, a = 0.025. B, a = 0.05. C, a = 0.25. In all three cases, critical values for the MF method are identical to those for UPGMA.

(not shown), the reason for this phenomenon is apparent. At the first step, the UPGMA and MF algorithms are identical—they join the two most similar taxa. If the length of b is increased beyond a certain threshold, then species A and C become the most similar pair of taxa, and they get joined at the first step, by both MF and UPGMA. Once this happens, the MF method never adds another taxon at a position that breaks up the original pair. A round of global branch swapping at the end of the MF method might correct this problem, but that was not examined in the present study.

The following three features are apparent in figure 2: (1) The sensitivity to rate variation changes considerably over relatively small changes in the overall rate of substitution, for both the P method and the NJ method, when incompletely corrected distances are used. (2) As expected, UPGMA is the method most sensitive to rate variation. UPGMA is quite robust for tree topologies such as the one examined in figure 2, which has a high stem-iness value (Fiala and Sokal 1985; Rohlf et al. 1990). Under these conditions, UPGMA can tolerate rate increases of twofold in species C (fig. 2B) or threefold in species B (fig. 2A). (3) The differences among the different methods are largest when the overall rate is low, but at those rates all the methods except UPGMA would be consistent even when rates in sister species differ by well over fivefold. When rates are high enough that inconsistency might become an issue, there is relatively little difference among the methods, although using a distance correction that is nearly exact will result in the NJ method being much more robust.

Generally the systematist has no control of the total λt values within the ingroup. Frequently, however, several choices of outgroup are available. The effect of keeping the branch lengths within the ingroup the same (a = 0.05) but choosing a progressively more distant outgroup is shown in figure 3. Again, when the length of branch b is increased (fig. 3A), the MF method behaves the same as UPGMA, but, when the length of branch c is increased (fig. 3B), MF tracks the NJ method. Clearly, an outgroup

Downloaded from https://academic.oup.com/mbe/article/9/3/537/1037266 by guest on 10 April 2024

that is as closely related to the ingroup as possible will maximize the robustness of any of the parsimony-based methods to varying rates. None of the methods will tolerate more than a 2.5-fold increase in the length of branch c when the outgroup is more distant [on the order of $\lambda t = 2.0 = 40a$ (fig. 3B)], although all of the methods except UPGMA can still tolerate a greater than fourfold increase in the length of branch b (fig. 3A).

The trees used to generate the results in figures 2 and 3 represent conditions where phylogeny inference should be relatively easy. The internal branches are the same length as branch a, resulting in a high stem-iness value (Fiala and Sokal 1985; Rohlf et al. 1990). It is reasonable to expect that, when the interior branches are shorter than a, all the methods will be much less robust to rate variation. Figure 4 presents results for three cases where there is nearly a trichotomy due to the very short length of branch f, which represents the common ancestor of species A and B. In figure 4A, the overall rate of change is relatively low, with a = 0.025. These particular λt values represent a crude approximation of one of the most notoriously difficult problems in systematics: the great-ape phylogeny [Nei et al. (1985) estimated that the total λt value for the human-gorilla pair is ~0.05 for the mitochondrial genome]. For figure 4A, the length of branch f is varied from 1% to 10% of the length of the terminal branch a (branches c–e are also varied along with f, to maintain the molecular clock). Under these conditions UPGMA is extremely sensitive to rate variation. When f = 0.01a, UPGMA is inconsistent, with as little as a 2% rate difference between species A and B, and, when f = 0.1a, UPGMA still becomes inconsistent, with only a 20% rate difference. The other methods perform better, and the P method can tolerate nearly a twofold difference in rate between A and B, even when f = 0.01a. In figure 4A, the slope of the line for the P method changes markedly at f/a = 0.07. When f < 0.07a and when b is above the line in figure 4A, P is inconsistent. In these cases, the shortest tree incorrectly has B as the sister species to (A,C). When f ≥ 0.7a and b is above the line in figure 4A, the P method incorrectly makes B the sister species to [(A,C),D]. Once this happens, the robustness of the P method to rate variation between branches A and B becomes nearly independent of the length of f. This effect also explains why the P method is more robust than the NJ method when the Jukes-Cantor correction is used when f/a is small—but less robust when f = a, as in figures 2A and 3A.

As expected, increasing the overall rate while keeping branch f short lowers the robustness of all the parsimony methods. The results for a = 0.05 are shown in figure 4B, and those for a = 0.25 in figure 4C. Again, UPGMA stays the same as the overall rate is increased, and the parsimony methods approach the sensitivity of UPGMA as the rates are increased. Particularly when a = 0.25, all the methods are quite sensitive to rate variation (note the different scales on the vertical axes in fig. 4). Even when f = 0.10a, the P and NJ methods (when Jukes-Cantor distances are used) can tolerate only ~35% difference in rate between species C and D.

Hendy and Penny (1989) discovered certain conditions under which the P method will be an inconsistent estimator of phylogeny even under a clocklike model of evolution. The particular case was a five-taxon problem with a symmetrical topology for the four members of the ingroup (made by rooting the tree in fig. 1 along branch c). The P method can be inconsistent when the internal branches (f and g) are very much shorter than the terminal branches. The results in figure 5 were obtained in the following way: For each point in figure 5, the total length of the tree is fixed, both from the root to the tips and from the first split in the ingroup to the tips (so a+f is constant). The

Downloaded from https://academic.oup.com/mbe/article/3/3/314/972609 by guest on 20 April 2024

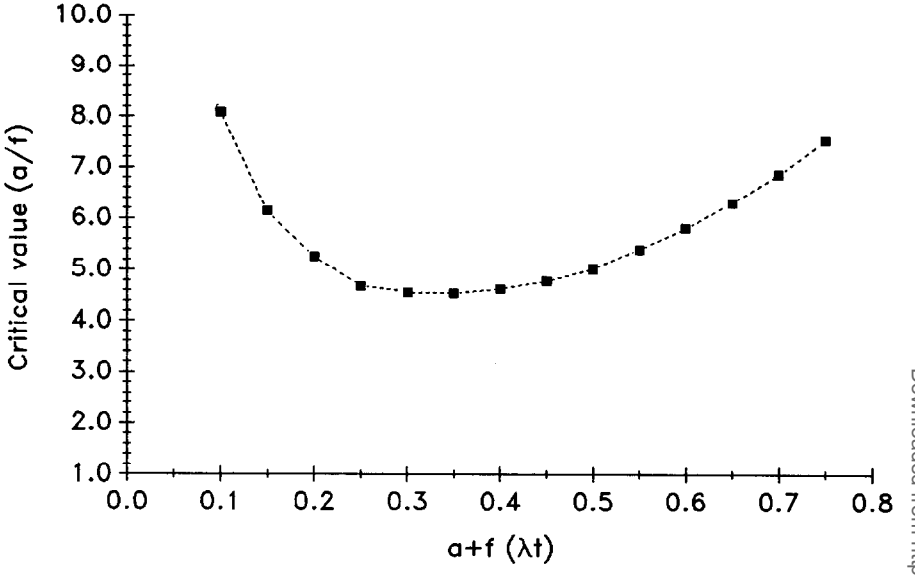


FIG. 5.—Relationship between consistency and ratio of length of internal branch f to length of terminal branch a for parsimony. A molecular clock holds for all branches. The tree is rooted along the branch leading to C in fig. 1, and the branch lengths are as given in table 1. See text for details.

first branching is halfway to the tips, so the total length of the outgroup branch is 3 × (a+f). While these values are held constant, the position of the second branchings (which produce the terminal taxa) are varied. The critical value is expressed as the ratio a/f; at the critical value, and for smaller f and larger a, the P method is inconsistent. For example, when a+f = 0.1 and when c = 0.3, then the P method is consistent when a = 0.088 and f = 0.012 but is inconsistent when a = 0.089 and f = 0.011. The critical value is 0.089/0.011, or ~8.1:1. Note that the molecular clock holds for all branches in this scenario. Hendy and Penny's (1989) result was surprising, as are the results in figure 5. The critical value is relatively high for short trees—and it falls as the length is increased, reaches a minimum, and then rises again as the length of the tree continues to increase. At the worst, the P method would be inconsistent when a is only ~4.5 times f. This disturbing behavior of the P method should be examined in more detail, particularly to determine whether these results apply to symmetrical subtrees of larger trees. Hendy and Penny's (1989) Hadamard matrix method for calculating expected character distributions would be much more efficient for examining larger trees than would the brute-force method used in the present study.

Discussion

By far the biggest difficulty in interpreting the results of the present study—and of studying the statistical properties of phylogenetic methods in general—is the enormous complexity of the parameter space. Even though the present study is limited to one of the simplest possible situations, only a few of the many parameters are varied, and only a limited range of values are tested. I have not examined cases where (*a*) more than one lineage departs from the molecular clock, (*b*) the rate of substitution varies among sites, or (*c*) sites are not independent, and so on. For this reason, caution must be used in interpreting these results. Still, the approach taken here, where the expected data-outcome probabilities are calculated directly, allows examination of

Downloaded from https://academic.oup.com/mbe/article/6/6/547/1033187/ by guest on 10 April 2024

much more of the parameter space than is possible by using simulations [e.g., see the study by Li et al. (1987), who were able to examine only a few sets of branch lengths].

It is not clear whether the results for five taxa can be extrapolated to larger problems, although it seems unlikely that the methods studied here will be more robust to violations of the molecular clock when there are more species involved. It is also not clear how these results relate to phylogenetic characters other than nucleotide sequences. For example, the evolution of morphological characters cannot be described by the sort of simple mathematical expressions used in the present study (even for nucleotides, those equations are, at best, a first-order approximation). Morphological characters are frequently analyzed by methods not examined in the present study, by using character states that are ordered into transformation series (meaning that some states can only be reached by first changing to some other state).

Felsenstein (1988) speculated that parsimony would probably be a consistent estimator of phylogeny under even a very rough molecular clock. A reasonable interpretation of the results in figures 2 and 3 is that all of the methods studied, including UPGMA, are quite robust to rate variation among taxa, over a broad range of overall rates, provided that the internal branches are not very short relative to the terminal branches. There are probably few situations where nucleotide substitution rates of sister taxa differ by a factor of ≥ 2. However, many real phylogenies will have some internal branches that are very short relative to the terminal branches. The results in figure 4 demonstrate that parsimony methods can be extremely sensitive to violations of the molecular clock when the overall rate of change is not very low and when an internal branch is short. I know of no methods that can determine whether the results of a particular phylogenetic analysis are incorrect because of a violation of the molecular clock. Relative-rate tests can be performed, but they will not be conclusive. There are situations where rates in sister taxa can greatly differ and where parsimony will be a consistent estimator. On the other hand, there are also situations where a large number of characters are required for detection of any rate differences but where parsimony is yet an inconsistent estimator. Further, proper use of a relative-rate test requires that the phylogeny used for the test be correct. If a high rate in one species results in an incorrect tree in which the offending species has moved toward the root, then the rates will appear much more uniform than they actually have been.

With regard to the distance-parsimony methods, some distance correction is better than none [provided that the data are not metric to begin with (Fitch 1980)]. The NJ method is more sensitive to rate variation when raw distances are used than it is when the Jukes-Cantor correction is applied to the distance matrix. This agrees with both Saitou (1988) and Saitou and Imanishi (1989), who found that using transformed distances resulted in obtaining the correct tree more often than was the case when the raw proportion of differing nucleotide sites was used.

In addition to the studies just mentioned, there have been a number of simulation studies that compare the "accuracy" of various methods (e.g., see Tateno et al. 1982; Saitou and Nei 1987; Sourdis and Krimbas 1987; Kim and Burgman 1988; Rohlf et al. 1990), where accuracy is usually defined as the proportion of replicated simulations for which the exactly correct tree was recovered. Accuracy is usually determined for several different sample sizes and is related to the statistical property of efficiency. Many of those studies examine cases where the rate of evolution varies among branches, and in such cases the estimation of efficiency is confounded by the fact that the consistency or inconsistency of each method has not been determined. In fact, all the complexities of the parameter space in the present study are present in the simulation

Downloaded from https://academic.oup.com/mbe/article/3/3/314/1037568 by guest on 19 April 2024

studies of accuracy—in addition to the new (but very real and very important) parameter of sample size. Further, the simulation studies generally use more than five taxa (eight taxa are commonly used), which makes the problem even more complex. Those results are very difficult to properly interpret in the absence of information about consistency.

It is tempting to use the results of the present study to draw conclusions about which method should be used. For example, as expected, UPGMA was always the method most sensitive to rate variation, and using uncorrected distances probably involves taking unnecessary risks. However, the differences found between the methods in the present study are not large, at least in the cases where any of the parsimony methods are so sensitive to rate variation that such differences might be a problem in real data. Further, which method would be most robust under a particular set of conditions would depend on how accurately the distances can be corrected. Therefore, the decision about which method to use should probably not be based simply on the fact that one is slightly more robust to rate variation than is another.

The present study examines the behavior of these methods if an infinitely large data set were available, but real data sets are actually rather small. The most important message from the simulation studies of accuracy is that the probability of obtaining the exactly correct tree from one reasonably sized data set is usually very low (e.g., see Rohlf et al. 1990). If we accept the view that phylogenetic analysis is actually a problem of statistical inference, then a tree derived from a single, relatively small data set is only an estimate made with error. It is entirely unrealistic to expect that a method be able to produce the single correct tree. Instead, we should be prepared to examine a range of trees, each of which is within the estimation error. In this regard, parsimony has a considerable advantage compared with the NJ and MF methods, since it is possible to assign a score to every topology and thus recognize which trees are the most nearly parsimonious. It is possible to use the parsimony score as a test statistic in this way, although the results of the present study confirm that the consistency of this estimate is by no means assured.

Downloaded from https://academic.oup.com/sysbio/article/9/3/537/1636760 by guest on 19 April 2024

Acknowledgments

I thank W. Fitch, J. Archie, and one anonymous reviewer for critical comments that have improved this paper, and I thank D. Straney, D. Swofford, D. Penny, L. Abele, and J. Felsenstein for helpful discussions over the years. I also thank W. Marzluff for providing space and other types of support and for allowing me to pursue this line of research. This work is supported in part by the Supercomputer Computations Research Institute, which is partially funded by U.S. Department of Energy contract DE-FC05-85ER25000, and by NIH postdoctoral fellowship 1F32GM13200. Computations were performed on a variety of platforms, including Decstation 5000, IBM RS/6000, and Cray Y/MP.

LITERATURE CITED

AOKI, K., Y. TATENO, and N. TAKAHATA. 1981. Estimating evolutionary distance from restriction maps of mitochondrial DNA with arbitrary G + C content. J. Mol. Evol. **18**:1–8.

CAVENDER, J. A. 1978. Taxonomy with confidence. Math. Biosci. **54**:217–229.

COLLESS, D. H. 1970. The phenogram as an estimate of phylogeny. Syst. Zool. **19**:352–362.

ECK, R. V., and M. O. DAYHOFF. 1966. Atlas of protein sequence and structure 1966. National Biomedical Research Foundation, Silver Spring, Md.

EDWARDS, A. W. F., and L. L. CAVALLI-SFORZA. 1964. The reconstruction of evolutionary

trees. Pp. 67–76 *in* V. H. HEYWOOD and J. MCNEILL, eds. Phenetic and phylogenetic classification. Systematics Association, London.

FARRIS, J. S. 1972. Estimating phylogenetic trees from distance matrices. Am. Nat. **106**:645–668.

FELSENSTEIN, J. 1973. Maximum likelihood and minimum steps methods for estimating evolutionary trees from data on discrete characters. Syst. Zool. **22**:240–249.

———. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. **27**:401–410.

———. 1979. Alternative methods of phylogenetic inference and their interrelationship. Syst. Zool. **28**:49–62.

———. 1983. Methods for inferring phylogenies: a statistical view, Pp. 315–334 *in* J. FELSENSTEIN, ed. Numerical taxonomy: proceedings of a NATO Advanced Study Institute. Springer, Berlin.

FELSENSTEIN, J. 1988. Phylogenies from molecular sequences: inference and reliability. Annu. Rev. Genet. **22**:521–565.

FIALA, K. L., and R. R. SOKAL. 1985. Factors determining the accuracy of cladogram estimation: evaluation using computer simulation. Evolution **39**:609–622.

FITCH, W. M. 1971. Toward defining the course of evolution: minimum change for a specific tree topology. Syst. Zool. **20**:406–416.

———. 1980. Estimating the total number of nucleotide substitutions since the common ancestor of a pair of homologous genes: comparison of several methods and three betahemoglobin messenger RNA's. J. Mol. Evol. **16**:153–209.

GOLDMAN, N. 1990. Maximum likelihood inference of phylogenetic trees, with special reference to a Poisson process model of DNA substitution and to parsimony analysis. Syst. Zool. **39**: 345–361.

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

JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–132 *in* H. N. MUNRO, ed. Mammalian protein metabolism. Academic Press, New York.

KIM, J., and M. A. BURGMAN. 1988. Accuracy of phylogenetic-estimation methods under unequal evolutionary rates. Evolution **42**:596–602.

KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.

LI, W.-H. 1986. Evolutionary change of restriction cleavage sites and phylogenetic inference. Genetics **113**:187–213.

LI, W.-H., K. H. WOLFE, J. SOURDIS, and P. M. SHARP. 1987. Reconstruction of phylogenetic trees and estimation of divergence times under nonconstant rates of evolution. Cold Spring Harbor Symp. Quant. Biol. **52**:847–856.

NEI, M., J. C. STEPHENS, and N. SAITOU. 1985. Methods for computing the standard errors of branching points in an evolutionary tree and their application to molecular data from humans and apes. Mol. Biol. Evol. **2**:66–85.

ROHLF, F. J., W. S. CHANG, R. R. SOKAL, and J. KIM. 1990. Accuracy of estimated phylogenies: effects of tree topology and evolutionary model. Evolution **45**:1671–1684.

ROHLF, F. J., and M. C. WOOTEN. 1988. Evaluation of the restricted maximum-likelihood method for estimating phylogenetic trees using simulated allele-frequency data. Evolution **43**:581–595.

SAITOU, N. 1988. Property and efficiency of the maximum likelihood method for molecular phylogeny. J. Mol. Evol. **27**:261–273.

SAITOU, N., and T. IMANISHI. 1989. Relative efficiencies of the Fitch-Margoliash, maximum-parsimony, maximum-likelihood, minimum-evolution, and neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. Mol. Biol. Evol. **6**:514–525.

Downloaded from https://academic.oup.com/mbe/article/9/3/537/1007266 by guest on 18 April 2024

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SNEATH, P. M. A., and R. R. SOKAL. 1973. Numerical taxonomy. W. H. FREEMAN, San Francisco.

SOURDIS, J., and C. KRIMBAS. 1987. Accuracy of phylogenetic trees estimated from DNA sequence data. Mol. Biol. Evol. **4**:159–166.

TATENO, Y., M. NEI, and F. TAJIMA. 1982. Accuracy of estimated phylogenetic trees from molecular data. I. Distantly related species. J. Mol. Evol. **18**:387–404.

WALTER M. FITCH, reviewing editor

Received March 22, 1991; revision received November 10, 1991

Accepted December 10, 1991

Downloaded from https://academic.oup.com/mbe/article/9/3/537/1037266 by guest on 10 April 2024