J Mol Evol (1988) 27:261–273

Journal of Molecular Evolution

Property and Efficiency of the Maximum Likelihood Method for Molecular Phylogeny

Naruya Saitou*

Center for Demographic and Population Genetics, The University of Texas Health Science Center at Houston, PO Box 20334, Houston, Texas 77225, USA

Summary. The maximum likelihood (ML) method for constructing phylogenetic trees (both rooted and unrooted trees) from DNA sequence data was studied. Although there is some theoretical problem in the comparison of ML values conditional for each topology, it is possible to make a heuristic argument to justify the method. Based on this argument, a new algorithm for estimating the ML tree is presented. It is shown that under the assumption of a constant rate of evolution, the ML method and UPGMA always give the same rooted tree for the case of three operational taxonomic units (OTUs). This also seems to hold approximately for the case with four OTUs. When we consider unrooted trees with the assumption of a varying rate of nucleotide substitution, the efficiency of the ML method in obtaining the correct tree is similar to those of the maximum parsimony method and distance methods. The ML method was applied to Brown et al.'s data, and the tree topology obtained was the same as that found by the maximum parsimony method, but it was different from those obtained by distance methods.

Key words: Molecular phylogeny — Maximum likelihood method — Tree-making methods — DNA sequence data

Offprint requests to: N. Saitou
* *Current address:* Department of Anthropology, Faculty of Science, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan

Introduction

The maximum likelihood (ML) method for constructing phylogenetic trees was first studied by Cavalli-Sforza and Edwards (1967) for the case of gene frequency data. Later, Felsenstein (1973, 1981) developed ML algorithms for constructing unrooted phylogenetic trees from amino acid or nucleotide sequence data. Kashyap and Subas (1974) also used the ML method for estimating a rooted tree from three amino acid sequences, assuming a constant rate of amino acid substitution. In Felsenstein's method, the ML value is computed for as many topologies as possible, and the topology that shows the highest ML value is chosen as the final tree. However, because of computational difficulty, the ML method is not used frequently. Furthermore, there are some theoretical problems that should be clarified before its application.

One problem is that the likelihood function to be used varies from topology to topology, so the ML values for different topologies are conditional and cannot be compared in the usual statistical sense (Nei 1987, pp. 323–325). Felsenstein (1984) tried to justify his algorithm by using Bayes' theorem, but his argument does not seem to be justified, because different likelihood functions require different probability spaces and we usually do not know the prior probability of each topology.

Nevertheless, it is likely that the ML value of the correct topology is generally higher than that of incorrect topologies. I have therefore studied statistical properties of the ML method applied for tree construction. As will be shown below, the ML value can indeed be used as a criterion for estimating the

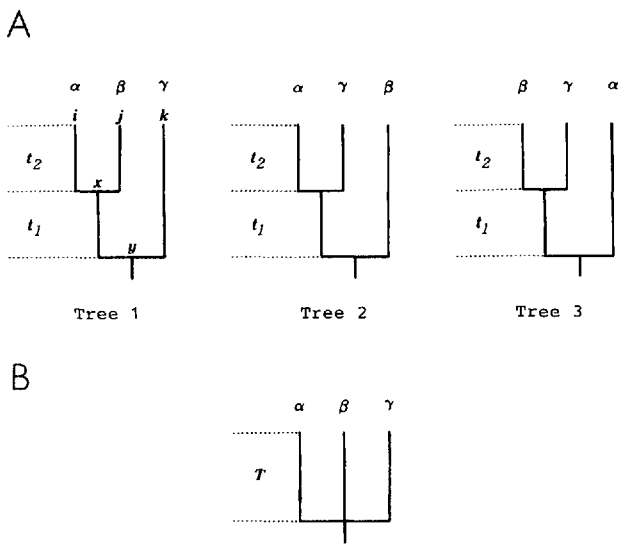

Fig. 1. **A** Three possible topologies for a rooted tree of three OTUs. **B** Trifurcation for a rooted tree of three OTUs.

correct topology and branch lengths. However, computer simulation has shown that the ML method does not show as good performance as some other tree-making methods in obtaining the correct tree at least under some situations considered in this study.

In the following, I shall consider the likelihood functions for rooted and unrooted trees separately and present a new algorithm for searching for the topology with the highest ML value. I shall also investigate the efficiency of the ML method in obtaining the correct tree in comparison with other tree-making methods.

Rooted Trees

In the case of rooted trees, we assume constancy of evolutionary rate. Thus, the number of parameters to be estimated for n operational taxonomic units (OTUs) is n − 1 for bifurcating rooted trees.

Three OTUs

Likelihood Function

Let us consider the simplest case, in which three OTUs are involved. There are three possible topologies (see Fig. 1A) and five different nucleotide configurations (see Table 1). Here, a nucleotide configuration means a particular pattern of nucleotide differences at a site among the three OTUs involved. The likelihood function, L(j), for the j-th topology (j = 1, 2, and 3) is given by

$$L(j) = U_{1j}^{m_1} U_{2j}^{m_2} U_{3j}^{m_3} U_{4j}^{m_4} U_{5j}^{m_5} \times C, \quad (1)$$

where U_{ij} is the probability of obtaining the i-th nucleotide configuration for the j-th topology, m_i is the observed number of the i-th nucleotide configuration, and $C = m!/(m_1!m_2!m_3!m_4!m_5!)$. m is the total number of nucleotides examined and is equal to the sum of m_i (i = 1, . . . , 5).

Table 1. Five possible nucleotide configurations for the case of three OTUs

Nucleotide configuration[a]					Probability for tree[b]		
No.	α	β	γ	Obs. no.	1	2	3
1	A	A	A	m_1	U_1	U_1	U_1
2	A	A	B	m_2	p	q	q
3	A	B	A	m_3	q	p	q
4	B	A	A	m_4	q	q	p
5	A	B	C	m_5	U_5	U_5	U_5

[a] α, β, and γ are OTUs of Fig. 1, and A, B, and C are nucleotides that are different from each other
[b] See Fig. 1A

To compute U_{ij}'s, we must know the pattern of nucleotide substitution and the structure (both the topology and branch lengths) of the tree under consideration. We first consider Jukes and Cantor's (1969) one-parameter model, in which a nucleotide changes to any other one with an equal probability. In this model, the probability that the nucleotide for a given site at time t is identical with that at time 0 is

$$P_{ii}(t) = \tfrac{1}{4} + \tfrac{3}{4} \exp(-\tfrac{4}{3}\lambda t), \quad (2a)$$

whereas the probability that the nucleotide i at time 0 changes to nucleotide j at time t is

$$P_{ij}(t) = \tfrac{1}{4} - \tfrac{1}{4} \exp(-\tfrac{4}{3}\lambda t), \quad (2b)$$

where λ is the rate of nucleotide substitution per nucleotide site. λ is assumed to be the same for all sites and is constant over all branches of a phylogenetic tree.

We now consider the probability of having nucleotides i, j, and k at a given nucleotide site for sequences α, β, and γ, respectively, under the assumption that tree 1 in Fig. 1A is the correct tree. This probability is given by

$$f(i, j, k) = \sum_y g_y \left\{ P_{yk}(t_1+t_2) \sum_x [P_{xy}(t_1) P_{xi}(t_2) P_{xj}(t_2)] \right\}, \quad (3)$$

where x and y are ancestral nucleotides (see Fig. 1) and g_y is the probability of having nucleotide y at this site. We assume that the nucleotide frequencies in DNA sequences are at equilibrium, so that g_y = 0.25 for all nucleotides in the one-parameter model.

Probabilities U_{i1} (i = 1, . . . , 5) for tree 1 in Fig.

1A can be obtained by applying equation (3) as follows.

$$U_{11} = 4f(i, i, i) = (A + Ba^2)/16, \quad (4a)$$
$$U_{21} = 12f(i, i, j) = (3A - Ba^2)/16, \quad (4b)$$
$$U_{31} = 12f(i, j, i) = (C + Da^2)/16, \quad (4c)$$
$$U_{41} = U_{31}, \quad (4d)$$
$$U_{51} = 24f(i, j, k) = (C - Da^2)/16, \quad (4e)$$

where $A = 1 + 3b^2$, $B = 6(1 + b)b^2$, $C = 3(1 - b^2)$, $D = 6(1 - b)b^2$, and $a = \exp(-4\lambda t_1/3)$ and $b = \exp(-4\lambda t_2/3)$. t_1 and t_2 are branch lengths shown in Fig. 1A. Substituting these quantities into equation (1), one can evaluate the likelihood, L(1), of tree 1 under specific t_1 and t_2 values. Computation of L(2) and L(3) is done in a similar manner.

If we assume Kimura's (1980) two-parameter model, in which transitions and transversions can occur at different rates, the equations that correspond to (2a) and (2b) are somewhat complicated (Li 1986; Saitou and Nei 1986). Consequently, f(i, j, k) in equation (3) should be modified, though the main structure of the formula is the same. There are many other substitution models, such as Takahata and Kimura's (1981) four-parameter model, the six-parameter model (Kimura 1981; Gojobori et al. 1982a), and the equal output model (Tajima and Nei 1984). It is straightforward to construct equations corresponding to (2) to (4) for each model of nucleotide substitution.

Analytical solution for the maximum likelihood (ML) value is not easy, but the ML value can be obtained numerically by changing t_1 and t_2. Note that U_{ij}'s are all functions of $v_1 \equiv \lambda t_1$ and $v_2 \equiv \lambda t_2$. Therefore, the ML value is obtainable by varying v_1 and v_2 numerically. As will be discussed later, the ML solutions of v_1 and v_2 are close to those obtained by the unweighted pair-group method (UPGMA; Sokal and Sneath 1963). Therefore, one may use these values as the initial values of v_1 and v_2.

Condition for Obtaining the Correct Tree

In the following we derive the condition for obtaining the correct tree for three OTUs, assuming that tree 1 of Fig. 1A is the correct tree. It will be shown that this condition is identical to that of UPGMA.

Let us assume that v_1 ($= \lambda t_1$) and v_2 ($= \lambda t_2$) are the same for all three topologies in Fig. 1A. Then these three topologies become identical if we neglect labels of the OTUs. Thus, $U_{11} = U_{12} = U_{13}$ ($= U_1$) and $U_{51} = U_{52} = U_{53}$ ($= U_5$). On the other hand, $U_{21} = U_{32} = U_{43}$ ($= p$), and similarly the remaining six U_{ij}'s are all the same and are designated as q (see Table 1). Therefore,

$$L(1) = U_1^{m_1} p^{m_2} q^{m_3+m_4} U_5^{m_5} \times C, \quad (5a)$$
$$L(2) = U_1^{m_1} p^{m_3} q^{m_2+m_4} U_5^{m_5} \times C, \quad (5b)$$
$$L(3) = U_1^{m_1} p^{m_4} q^{m_2+m_3} U_5^{m_5} \times C. \quad (5c)$$

We take the logarithm of L(j)s (log-likelihoods) for computational convenience. For example,

$$\log L(1) = m_1 \log U_1 + m_2 \log p + (m_3 + m_4)\log q + m_5 \log U_5 + \text{const.} \quad (6)$$

Let us now compare the log-likelihood of tree 1 with those of trees 2 and 3. We then have

$$\log L(1) - \log L(2) = (m_2 - m_3)\cdot(\log p - \log q), \quad (7a)$$
$$\log L(1) - \log L(3) = (m_2 - m_4)\cdot(\log p - \log q). \quad (7b)$$

From equations (4b) and (4c), we can see that

$$p - q = U_{21} - U_{31} = 3b^2(1 + a)(1 - a)/4. \quad (8)$$

Since $0 < a < 1$ for $t_1 > 0$, $p - q > 0$ and $\log p - \log q$ in equations (7a) and (7b) is always positive. Thus, L(1) > L(2) and L(1) > L(3) if

$$m_2 > m_3 \quad \text{and} \quad m_2 > m_4. \quad (9)$$

Because this is true for any set of t_1 and t_2, the ML value for tree 1 must be larger than those for trees 2 and 3 if inequalities (9) hold. Therefore, (9) is the condition to obtain the correct tree by the ML method. Interestingly, this condition is the same as that for UPGMA (Saitou and Nei 1986). This means that the topology of the UPGMA tree is always identical with that of the maximum likelihood tree, though the branch lengths (v_1 and v_2) may be different.

Numerical Examples

Figure 2A is an example of likelihood surfaces for three topologies. Here, tree 1 of Fig. 1A is assumed to be the correct tree, and $v_1 = 0.04$, $v_2 = 0.06$, $U_{11} = 0.7746$, $U_{21} = 0.1145$, $U_{31} = U_{41} = 0.0498$, and $U_{51} = 0.0113$ are used. The m_i values were determined by using pseudorandom numbers, assuming $m = 1000$. m_i's thus obtained were $m_1 = 789$, $m_2 = 98$, $m_3 = 59$, $m_4 = 50$, and $m_5 = 4$. Since $m_2 > m_3$ and $m_2 > m_4$ in this example, tree 1 gives the highest maximum likelihood value. Numerical evaluation of the likelihood surface gives a maximum log-likelihood value of -756.36 with $\hat{v}_1 = 0.0259$ and $\hat{v}_2 = 0.0606$. Figure 2A shows the likelihood surfaces for three trees. These curves represent the ML values for given v_1 values (an ML value was computed by varying v_2 for a given v_1 value). It is clear that the maximum likelihood estimates of v_1 for trees 2 and 3 are both zero (in this case, the log-likelihood value of -765.32 was obtained with $\hat{v}_2 = 0.0776$). That is, the trifurcating tree shown

in Fig. 1B gives the maximum likelihood solution for these two erroneous topologies. This can be explained by considering the branch length estimates by a distance method, as shown below.

In the above example, the number of nucleotide differences (n_{ij}) for OTUs i and j are $n_{\alpha\beta} = m_3 + m_4 + m_5 = 113$, $n_{\alpha\gamma} = m_2 + m_4 + m_5 = 152$, and $n_{\beta\gamma} = m_2 + m_3 + m_5 = 161$. The Jukes-Cantor distance (d_{ij}) between OTUs i and j is given by

$$d_{ij} = -\tfrac{3}{4} \log\left(1 - \tfrac{4}{3}\frac{n_{ij}}{m}\right), \tag{10}$$

where m (= 1000) is the total number of nucleotides compared. Therefore, we have $d_{\alpha\beta} = 0.1225$, $d_{\alpha\gamma} = 0.1699$, and $d_{\beta\gamma} = 0.1812$. If we use UPGMA, tree 1 is chosen, and branch length estimates become $\hat{v}_1 = (d_{\alpha\gamma} + d_{\beta\gamma})/4 - d_{\alpha\beta}/2 = 0.0265$ and $\hat{v}_2 = d_{\alpha\beta}/2 = 0.0613$. These are close to the ML estimates. The log-likelihood corresponding to these estimates is -756.38, which is again very close to the maximum log-likelihood value. When tree 2 is considered, however, we have to cluster OTUs α and γ first. If we use a method of estimating branch lengths similar to that of UPGMA, $\hat{v}_2 = d_{\alpha\gamma}/2 = 0.0850$ and $\hat{v}_1 = (d_{\alpha\beta} + d_{\beta\gamma})/4 - \hat{v}_2 = -0.0091$ are obtained. Apparently because the estimate for v_1 becomes negative by the distance method, the trifurcation ($v_1 = 0$) gives the best fit for tree 2 when the ML method is used. The same thing can be shown for tree 3.

Under a certain condition, however, two of the three possible trees may have positive v_1 estimates. Let us assume $m_2 > m_3 > m_4$. We then have

$$d_{\alpha\beta} < d_{\alpha\gamma} < d_{\beta\gamma}. \tag{11}$$

Under this condition, tree 1 of Fig. 1A is chosen with a positive estimate for v_1 either by UPGMA or by the ML method. Let us further assume that

$$d_{\alpha\gamma} < (d_{\alpha\beta} + d_{\beta\gamma})/2. \tag{12}$$

In this case, the distance method gives a positive estimate of v_1 when tree 2 is considered, and the ML value for tree 2 is higher than that for the trifurcating tree. An example is given in Fig. 2B, where the case of m = 100, $m_1 = 40$, $m_2 = 20$, $m_3 = 15$, $m_4 = 5$, and $m_5 = 20$ is considered. Therefore, we have $d_{\alpha\beta} = 0.5716$, $d_{\alpha\gamma} = 0.6872$, and $d_{\beta\gamma} = 0.9913$. Since $m_2 > m_3 > m_4$, tree 1 of Fig. 1A gives the highest log-ML value (-151.00 with $\hat{v}_1 = 0.1362$ and $\hat{v}_2 = 0.2947$), which is higher than that for the trifurcating tree (log-likelihood = -152.90 with $\hat{v}_1 = 0$ and $\hat{v}_2 = 0.3799$). However, tree 2 also has an ML value (log-likelihood = -152.77 with $\hat{v}_1 = 0.0355$ and $\hat{v}_2 = 0.3566$) higher than that for the trifurcating tree. UPGMA chooses tree 1 with $\hat{v}_1 = 0.1338$ and $\hat{v}_2 = 0.2858$. But if we cluster OTUs α and γ (tree 2), we have $\hat{v}_1 = 0.0471$ and $\hat{v}_2 = 0.3436$. Thus, the estimate for v_1 is positive. This corresponds to the case where the ML value for tree 2 is higher than that for the trifurcating tree. For tree 3, however, $\hat{v}_1$ by the distance method is negative, and the ML value decreases as the absolute value of v_1 increases (see Fig. 2B).

Four OTUs

Likelihood Function

When there are four OTUs, the situation becomes much more complicated. The number of possible bifurcating trees is now 15, and 3 of them (trees 3a, 3b, and 3c) are shown in Fig. 3. We have to consider 15 nucleotide configurations (see Table 2), and the likelihood function varies with tree topology. For example, the likelihood function for tree 3a is given by

$$L(3a) = \prod_{i=1}^{15} U_i^{m_i}, \tag{13}$$

where U_i is the probability of having the i-th nucleotide configuration and m_i is the observed number of the i-th configuration. U_i is computed in a manner similar to the case of three OTUs by using a function below.

$$f(i, j, k, l) = \sum_x g_x P_{xl}(t_1 + t_2 + t_3) \sum_y \left\{ P_{xy}(t_1) P_{yk}(t_2 + t_3) P_{yz}(t_2) \sum_z [P_{zi}(t_3) P_{zj}(t_3)] \right\}, \tag{14}$$

where i, j, k, and l are the nucleotides observed in OTUs α, β, γ, and δ, respectively, and x, y, and z are the nucleotides at the three ancestral nodes (see tree 3a of Fig. 3). g_x is the probability of having nucleotide x at this site. We assume that the nucleotide frequencies in DNA sequences are at equilibrium, so that $g_x = 0.25$ for all four nucleotides. Thus, for example, U_1 for tree 3a of Fig. 3 becomes 4f(i, i, i, i).

Condition for Obtaining the Correct Tree

Comparison of trees 3a to 3c is similar to the case of three OTUs discussed in the previous section. Although the ML method and UPGMA may no longer give the same topology, these two methods are expected to give similar results. Table 2 presents the 15 nucleotide configurations. If we use the same v_1, v_2, and v_3 ($v_i \equiv \lambda t_i$, where λ is the rate of nucleotide substitution) values for trees 3a, 3b, and 3c of Fig. 3, the probability of obtaining 15 configurations can be described by 11 probabilities (see

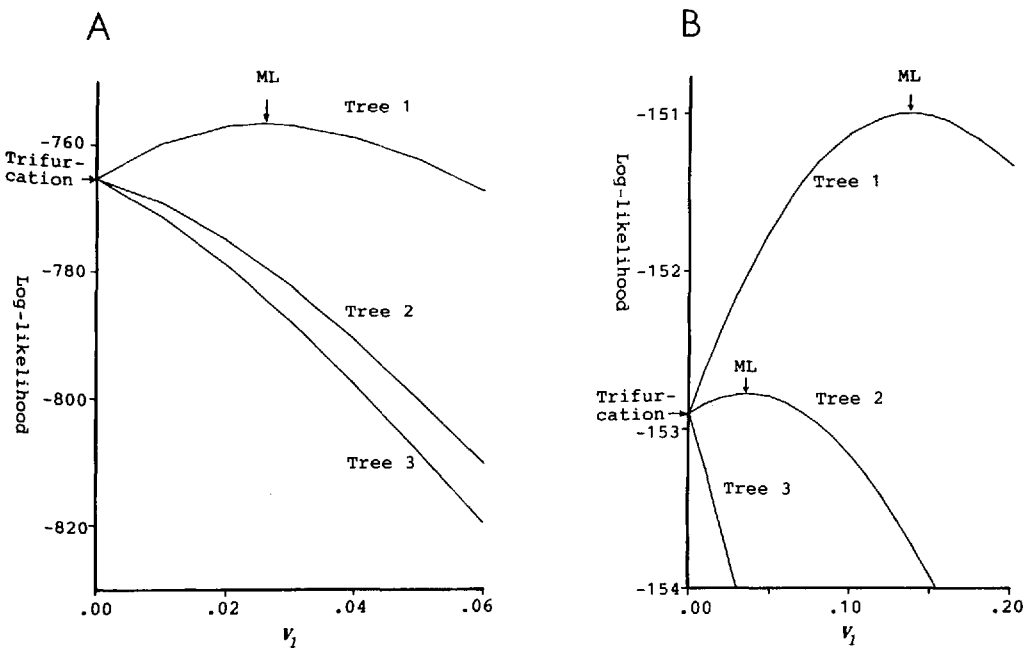

Fig. 2. Two examples of likelihood surfaces for three topologies in Fig. 1A. **A** The case of $m_1 = 789$, $m_2 = 98$, $m_3 = 59$, $m_4 = 50$, and $m_5 = 4$. **B** The case of $m_1 = 40$, $m_2 = 20$, $m_3 = 15$, $m_4 = 5$, and $m_5 = 20$.

Table 2), and the differences of log-likelihood among these three trees can be written as

$$\begin{aligned} &L(3a) - L(3b) \\ &\quad = (m_3 - m_4)(\log a - \log b) \\ &\qquad + (m_6 - m_7)(\log c - \log d) \\ &\qquad + (m_9 - m_{10})(\log e - \log f) \\ &\qquad + (m_{12} - m_{13})(\log g - \log h), \end{aligned} \tag{15a}$$

$$\begin{aligned} &L(3a) - L(3c) \\ &\quad = (m_3 - m_5)(\log a - \log b) \\ &\qquad + (m_6 - m_8)(\log c - \log d) \\ &\qquad + (m_9 - m_{11})(\log e - \log f) \\ &\qquad + (m_{12} - m_{14})(\log g - \log h), \end{aligned} \tag{15b}$$

where L(i) is the likelihood for tree i. As in the case of $\log p - \log q$ for three OTUs, it can be shown that $\log a - \log b$, $\log c - \log d$, $\log e - \log f$, and $\log g - \log h$ are all positive. Therefore, $L(3a) > L(3b)$ and $L(3a) > L(3c)$ if

$$\begin{aligned} m_3 > m_4 \quad &\text{and} \quad m_3 > m_5, \\ m_6 > m_7 \quad &\text{and} \quad m_6 > m_8, \\ m_9 > m_{10} \quad &\text{and} \quad m_9 > m_{11}, \\ m_{12} > m_{13} \quad &\text{and} \quad m_{12} > m_{14}. \end{aligned} \tag{16}$$

UPGMA also chooses tree 3a when these inequalities hold, because the condition of obtaining tree 3a from three trees (3a, 3b, and 3c) by UPGMA is

$$\begin{aligned} &m_3 + m_6 + m_9 + m_{12} \\ &\quad > m_4 + m_7 + m_{10} + m_{13} \quad \text{and} \\ &m_3 + m_6 + m_9 + m_{12} \\ &\quad > m_5 + m_8 + m_{11} + m_{14}. \end{aligned} \tag{17}$$

Unless the amount of DNA divergence is large, $m_3, \ldots,$ and m_8 (two nucleotides are observed at these configurations) are usually larger than $m_9, \ldots,$ and m_{14} (three nucleotides are observed at these configurations). Similarly, probabilities a, b, c, and d are usually larger than e, f, g, and h. Therefore, inequality (16) can be approximated by

$$\begin{aligned} m_3 > m_4 \quad &\text{and} \quad m_3 > m_5, \\ m_6 > m_7 \quad &\text{and} \quad m_6 > m_8, \end{aligned} \tag{18}$$

and inequality (17) by

$$\begin{aligned} &m_3 + m_6 > m_4 + m_7 \quad \text{and} \\ &m_3 + m_6 > m_5 + m_8. \end{aligned} \tag{19}$$

Thus we expect that the topology estimated by the ML method is similar to that obtained by UPGMA, as in the case of three OTUs. We note that the branch length estimates of UPGMA are least squares estimates under the assumption of rate constancy for a given tree (Chakraborty 1977).

Algorithm for Finding the ML Tree

There are two levels of null or consensus trees for the case of four OTUs. One is for quadrifurcation (level I of Fig. 3) and the other is for one trifurcation and one bifurcation (level II of Fig. 3). The number of null trees is 1 for level I, but 10 for level II. Three of these 10 trees at level II are shown in Fig. 3. Each tree at level II produces three bifurcating trees (level III) when the trifurcation is resolved (e.g., trees 3a, 3b, and 3c from tree 2a). Therefore, one tree at level III is related to two trees at level II. For example, tree 3a is related to trees 2a and 2b in Fig. 3.

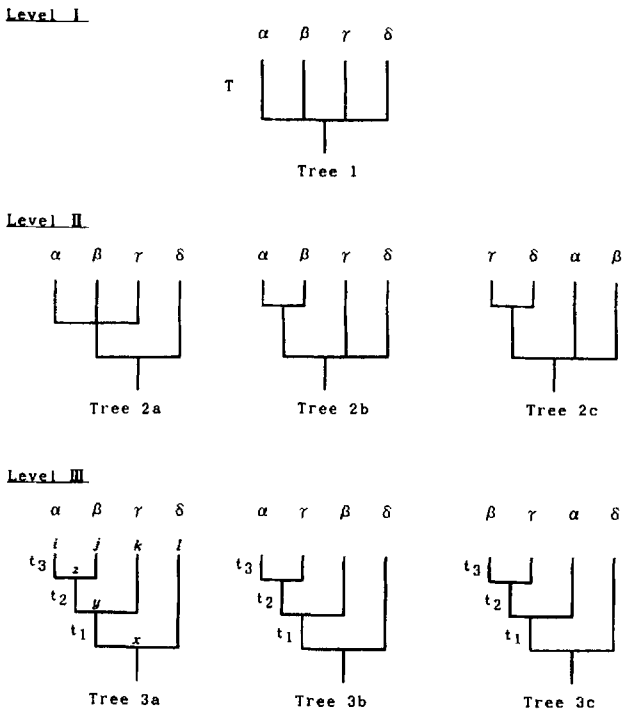

Fig. 3. Three levels of rooted trees for four OTUs

Table 2. Fifteen nucleotide configurations for the case of four OTUs

Nucleotide configuration[a]						Probability for tree[b]		
No.	α	β	γ	δ	Obs. no.	3a	3b	3c
1	A	A	A	A	m_1	U_1	U_1	U_1
2	A	A	A	B	m_2	U_2	U_2	U_2
3	A	A	B	A	m_3	a	b	b
4	A	B	A	A	m_4	b	a	b
5	B	A	A	A	m_5	b	b	a
6	A	A	B	B	m_6	c	d	d
7	A	B	A	B	m_7	d	c	d
8	A	B	B	A	m_8	d	d	c
9	A	A	B	C	m_9	e	f	f
10	A	B	A	C	m_{10}	f	e	f
11	B	A	A	C	m_{11}	f	f	e
12	B	C	A	A	m_{12}	g	h	h
13	B	A	C	A	m_{13}	h	g	h
14	A	B	C	A	m_{14}	h	h	g
15	A	B	C	D	m_{15}	U_{15}	U_{15}	U_{15}

[a] α, β, γ, and δ are OTUs of Fig. 3, and A, B, C, and D are different nucleotides

[b] See Fig. 3 for the designation of tree topologies

The two trees (2a and 2b) related to tree 3a at level III can be used as null trees to test the significance of t_1 and t_2 of tree 3a, respectively. If t_1 is not significantly greater than zero, we assume that tree 2a is the correct one, and proceed to test the significance of t_2 with tree 1 as the null tree. In general, there are n − 1 levels of trees with n OTUs, and we can successively test the significance of interior branches of a tree.

In the above, we implicitly assumed that the tree topology (one at level III) is known. However, the reverse process (descending from level I) may be used to find the final topology. This algorithm is reminiscent of the neighbor-joining method (Saitou and Nei 1987). Let us explain the procedure by using Fig. 3. First the ML value for the tree at level I (tree 1) is computed. The ML values for all 10 trees at level II are then computed and they are compared with each other. We choose the tree with the highest ML value among these 10 trees. Suppose it is tree 2a. We then compute the ML values for trees 3a, 3b, and 3c, and the tree with the highest ML value is chosen as the final tree.

The rationale for this algorithm is as follows. Suppose that tree 3a is the final solution with the highest conditional ML value. Trees 2a and 2b are then expected to have higher ML values than that of tree 1 or than those of other trees at level II. Though it is difficult to prove this statement analytically, numerical computations (not shown) do give the expected pattern. Interestingly, seven topologies that are not shown in Fig. 3 almost always have lower ML values than that of tree 1. This property is similar to that of the case of three OTUs. Note that tree 2c may have a higher ML value than that of tree 1. Interestingly, this tree and tree 2b become the same unrooted tree if we ignore the root. In any case, either tree 2a or 2b may show the highest conditional ML value, depending on the observed data. The comparisons at level III are then restricted to trees 3a, 3b, and 3c. In this way, we may be able to find the tree with the highest conditional ML value. When the number of OTUs is large, this algorithm requires much less computational time compared with the ordinary algorithm in which as many bifurcating trees as possible are examined.

Unrooted Trees

The procedure of the ML method for unrooted trees is somewhat different from that for rooted trees, in which constancy of evolutionary rate is assumed. Following Felsenstein (1981), we assume that the rate of nucleotide substitution varies from branch to branch, so that we need one parameter for each branch ($v_i \equiv \lambda_i t_i$, where λ_i and t_i are the values of λ and t for the i-th branch, respectively). The number of parameters to be estimated for n OTUs is n − 1 for bifurcating rooted trees, but we need 2n − 3 parameters for bifurcating unrooted trees. Because there is only one topology for the case of three OTUs, we start from the case for four OTUs.

Four OTUs

Likelihood Function

There are three possible unrooted trees for the case of four OTUs (see Fig. 4A), and we have to consider 15 nucleotide configurations as in the case of rooted trees. The likelihood function of tree j (j = 1, 2, and 3) of Fig. 4A is

$$L(j) = \prod_{i=1}^{15} U_{ij}^{m_i}, \qquad (20)$$

where U_{ij} is the probability for the i-th nucleotide configuration for the j-th tree and m_i is the observed number of the i-th nucleotide configuration (see Table 2). These U_{ij}'s are computed in the same way as those for the case of rooted trees. For example, the probabilities for tree 1 of Fig. 4A ($U_i \equiv U_{i1}$) are given by

$$\begin{aligned}
U_1 &= 4h(i, i, i, i), \\
U_2 &= 12h(i, i, i, j), \quad (i \neq j) \\
U_3 &= 12h(i, i, j, i), \quad (i \neq j) \\
U_4 &= 12h(i, j, i, i), \quad (i \neq j) \\
U_5 &= 12h(j, i, i, i), \quad (i \neq j) \\
U_6 &= 12h(i, i, j, j), \quad (i \neq j) \\
U_7 &= 12h(i, j, i, j), \quad (i \neq j) \\
U_8 &= 12h(i, j, j, i), \quad (i \neq j) \\
U_9 &= 24h(i, i, j, k), \quad (i \neq j \neq k) \\
U_{10} &= 24h(i, j, i, k), \quad (i \neq j \neq k) \\
U_{11} &= 24h(j, i, i, k), \quad (i \neq j \neq k) \\
U_{12} &= 24h(j, k, i, i), \quad (i \neq j \neq k) \\
U_{13} &= 24h(j, i, k, i), \quad (i \neq j \neq k) \\
U_{14} &= 24h(i, j, k, i), \quad (i \neq j \neq k) \\
U_{15} &= 24h(i, j, k, l), \quad (i \neq j \neq k \neq l) \qquad (21)
\end{aligned}$$

where

$$h(i, j, k, l) = \sum_x g_x \cdot \left\{ P_{xk}(v_3) P_{xl}(v_4) \cdot \left[\sum_y P_{yi}(v_1) P_{yj}(v_2) P_{xy}(v_5) \right] \right\}. \qquad (22)$$

Here, i, j, k, and l are the nucleotides at OTUs 1, 2, 3, and 4 of tree 1 of Fig. 4A, respectively, and the interior node (nucleotide x) that connects OTUs 3 and 4 is assumed to be the ancestor. In general, however, any point can become the ancestral one. This is the so-called pulley principle (Felsenstein 1981). g_x is the probability of having nucleotide x at a site, and we assume it to be 0.25 for all four nucleotides as before. v_i (i = 1, 2, 3, and 4) is the branch length between OTU i and its nearest interior node (either x or y), and v_5 is the length of the interior branch.

Computational Procedure

We first compute the ML value for the quadrifurcating tree (Fig. 4B), assuming $v_5 = 0$. This tree corresponds to the tree at level I. Branch length estimates obtained by the neighbor-joining method (Saitou and Nei 1987) may be used as the initial values of v_i's (i = 1, 2, 3, and 4). That is,

$$\hat{v}_i = \sum_{j=1}^{4} d_{ij} - \tfrac{2}{3} \sum_{j<k} d_{jk}, \qquad (23)$$

It can be shown that $\hat{v}_i$'s in equation (23) give the least squares estimate for the quadrifurcating tree of Fig. 4B (see Saitou and Nei 1987). After the maximum likelihood solution is numerically determined for the quadrifurcating tree, three trees in Fig. 4A are examined if a positive v_5 value can increase the ML value. An example of the computation is shown in Fig. 5. This is a result based on one replication of simulation. The parameters used are $v_1 = 0.1$, $v_2 = 0.4$, $v_3 = 0.2$, $v_4 = 0.3$, $v_5 = 0.05$, and 500 nucleotides are compared. The Jukes-Cantor model of nucleotide substitution is used both for generating sequence data and for the ML estimation. Using tree 1 as the model tree, the following 15 m_i values are obtained for i = 1, . . . , and 15: 196, 67, 40, 77, 22, 12, 5, 8, 11, 21, 10, 12, 3, 15, and 1. Tree 1 has the highest ML value (log-likelihood is −1004.2) when $\hat{v}_5 = 0.026$ (see Fig. 5). The estimates of other branch lengths are $\hat{v}_1 = 0.110$, $\hat{v}_2 = 0.361$, $\hat{v}_3 = 0.188$, and $\hat{v}_4 = 0.303$. On the other hand, the same topology is obtained when we use the neighbor-joining method, with $\hat{v}_1 = 0.101$, $\hat{v}_2 = 0.366$, $\hat{v}_3 = 0.179$, $\hat{v}_4 = 0.316$, and $\hat{v}_5 = 0.033$, which are close to the estimates obtained by the ML method. The corresponding log-likelihood value for these estimates is −1004.5, slightly lower than that of the ML estimate.

The maximum likelihood value for the other two unrooted trees is obtained for the case with no interior branch (the starlike tree in Fig. 4B). This situation is similar to that of Fig. 2 for the case of rooted, three OTUs. As shown in the previous section, however, other tree topologies may have ML values with a positive v_5 value.

Five OTUs

Likelihood Function

As in the case of rooted trees for four OTUs, there are 15 unrooted trees for five OTUs. The number of possible nucleotide configurations is now 51 (Saitou and Nei 1986), and the likelihood function for a tree is

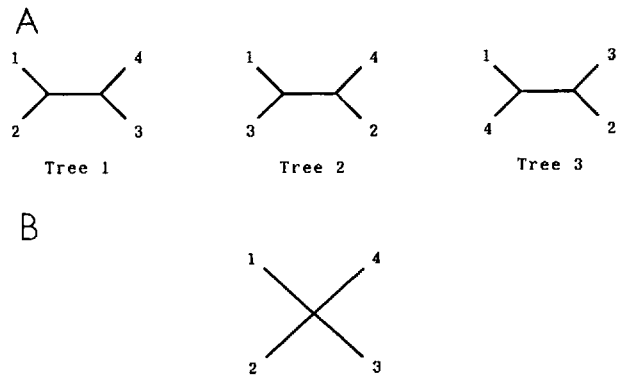

Fig. 4. **A** Three possible topologies for an unrooted tree of four OTUs. **B** Starlike tree for four OTUs.

$$L = \prod_{i=1}^{51} U_i^{m_i}, \tag{24}$$

where U_i is the probability for observing the i-th nucleotide configuration and m_i is the observed number of the i-th configuration. U_i's are determined in the same way as that for the case of four OTUs by using the following probability:

$$h(i, j, k, l, m) = \sum_x g_x \cdot \left\{ P_{xk}(v_3) \left[\sum_y P_{xy}(v_6) P_{yi}(v_1) P_{yj}(v_2) \right] \cdot \left[\sum_z P_{xz}(v_7) P_{zl}(v_4) P_{zm}(v_5) \right] \right\}, \tag{25}$$

where i, j, k, l, and m are nucleotides at OTUs 1, . . . , 5, respectively, and x, y, and z are nucleotides at three interior nodes (see Fig. 6). $g_x = 0.25$ as before.

Algorithm for Finding the ML Tree

As in the case of rooted trees, we first compute the ML estimate for the starlike tree (level I of Fig. 7) by setting the lengths of two interior branches to zero. The initial values for v_i's are computed from the distance matrix as

$$\hat{v}_i = \sum_{j=1}^{5} d_{ij} - \tfrac{3}{4} \sum_{j<k} d_{jk}. \tag{26}$$

Ten different trees are then considered, and the ML value for each topology is computed. These trees have one trifurcation and one bifurcation, and they correspond to level II trees (among them, three trees are shown in Fig. 7). The ML estimates of v_i's (i = 1, . . . , 5) for the starlike tree are used as the initial values. From each tree at level II, three trees are produced if the trifurcation is resolved (level III).

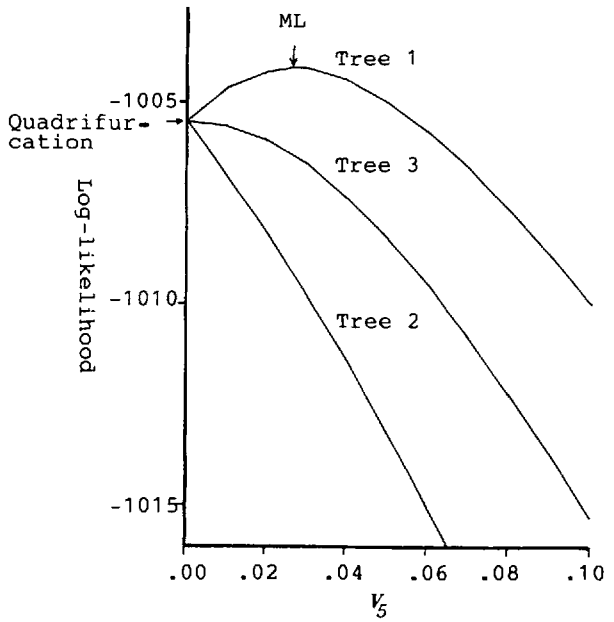

Fig. 5. An example of likelihood surfaces for three unrooted trees of four OTUs

For example, trees 3a, 3b, and 3c are produced from tree 2a in Fig. 7. The ML estimates of six branch lengths for the case of level II, which is chosen by the criterion of the highest ML value, are used as the initial values for the computation of the ML estimation for a tree at level III. In this case, the seven-dimensional likelihood surface is numerically examined for three possible trees. This new algorithm can be extended to any number of OTUs, as in the case for rooted trees.

Numerical Examples

Brown et al. (1982) determined sequences of 895 nucleotides of a portion of mitochondrial DNA from five hominoid species (humans, chimpanzees, gorillas, orangutans, and gibbons). We used these data as the example for constructing an unrooted tree of five OTUs by the maximum likelihood method developed above. Out of 51 possible configurations, 28 configurations were observed in Brown et al.'s (1982) data, and they are listed in Table 3.

The estimates of v_i's obtained by the distance method for the level I tree (tree a of Fig. 8) are 0.0522, 0.0617, 0.0653, 0.1144, and 0.1694 for the branch leading to humans, chimpanzees, gorillas, orangutans, and gibbons, respectively. Corresponding branch length estimates by the ML method are 0.0470, 0.0617, 0.0719, 0.1373, and 0.1694, respectively. The maximum log-likelihood value is −1409. Among 10 trees at level II, tree b, which clusters orangutans and gibbons, is selected because this tree has the highest log-likelihood value (−1359).

We now restrict the search of tree topology at level III to the three trees in which orangutans and

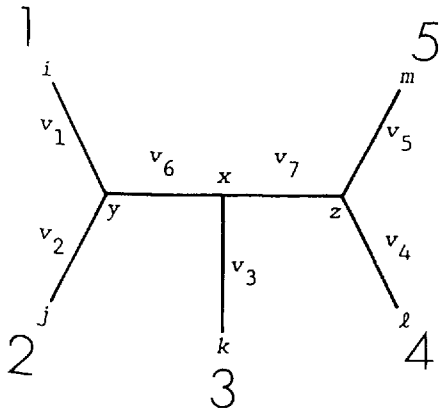

Fig. 6. An unrooted tree for five OTUs

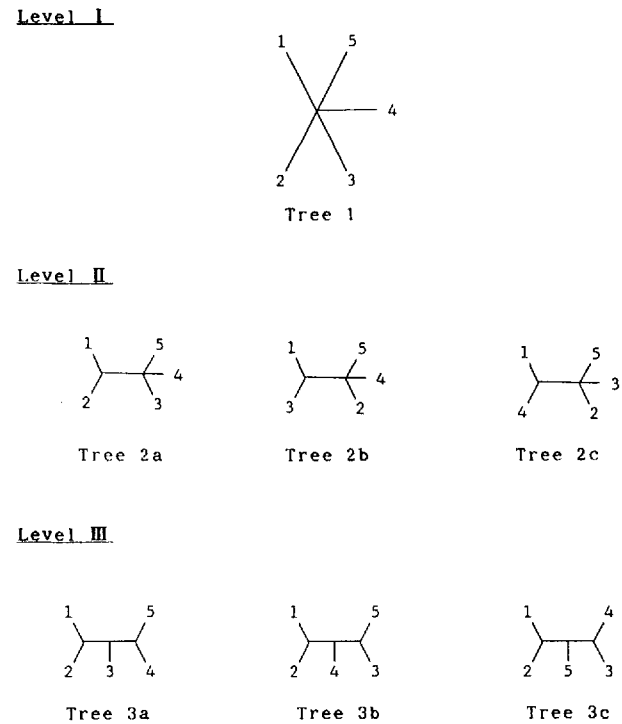

Fig. 7. Three levels of unrooted trees for five OTUs

gibbons are clusterd. Tree c1, which clusters chimpanzees and gorillas, is selected because it has the highest log-likelihood value (−1346) among the three trees. The maximum parsimony method chooses the same topology (Brown et al. 1982). The other two trees showed ML values of −1347 (tree c2; humans and chimpanzees are clustered) and −1354 (tree c3; humans and gorillas are clustered). If we apply UPGMA, tree c2 is chosen (Nei et al. 1985). Fitch and Margoliash's (1967) method and the distance Wagner method (Farris 1972) also choose tree c2 (Nei 1987, pp. 298–308); so do the transformed distance method (Farris 1977; Klotz and Blanken 1981; Li 1981) and the neighbor-joining method (Saitou and Nei 1987). The ML value for this tree is very close to the highest one (tree c1). Estimates of branch lengths for tree c2 are similar to those obtained by Fitch and Margoliash's (1967) method (see Nei 1987, p. 301). Note that Fitch and Margoliash's method gives the least squares estimate for a tree of five OTUs (N. Saitou, unpublished result). Hasegawa et al. (1985) applied J. Felsenstein's program [program DNAML in PHYLIP, based on Felsenstein (1981)] to Brown et al.'s (1982) data, and they found the highest ML value for tree c2. This occurred probably because they discarded transitional changes, and two other sequences (bovine and mouse) were included in their analysis.

Efficiency of the Maximum Likelihood Method

In this paper I studied the mathematical properties of the maximum likelihood method to some extent. However, the efficiency of a tree-making method can only be studied by simulation. Therefore, I present some results of computer simulations in which the ML method for unrooted trees is compared with other tree-making methods.

For trees of four OTUs, the efficiency of the ML method was compared with that of two other tree-making methods. One is the maximum parsimony method (Eck and Dayhoff 1966; Fitch 1977). Note that the compatibility method (Le Quesne 1969) and the maximum parsimony method are identical for four OTUs (Saitou and Nei 1986). The other tree-making method used is a distance method. Saitou and Nei (1986) showed that the condition to obtain the correct tree for four OTUs is the same for the distance Wagner method (Farris 1972), modified Farris methods (Tateno et al. 1982; Faith 1985), and transformed distance methods (Farris 1977; Klotz and Blanken 1981; Li 1981). The same condition also holds for neighborliness methods (Sattath and Tversky 1977; Fitch 1981) and the neighbor-joining method (Saitou and Nei 1987). For tree 1 of Fig. 4A, this condition is given by

$$d_{12} + d_{34} < d_{13} + d_{24},$$
$$d_{12} + d_{34} < d_{14} + d_{23}. \qquad (27)$$

Both the proportion of different nucleotides and the Jukes-Cantor distances are used for d_{ij}.

Table 4 shows the result when the model tree A of Fig. 9 is used. This model tree is the same as that of Fig. 3A of Saitou and Nei (1986). A constant rate of evolution is assumed in this model tree. Five hundred nucleotides were compared and 100 replications were obtained. For the distance method, the results for the proportion of different nucleotides are shown, since they performed slightly better than

Table 3. Observed nucleotide configurations of hominoid mitochondrial DNA data

Configuration[a]						No. of nucleotides observed
No.	Hu	Ch	Go	Or	Gi	
1	A	A	A	A	A	613
2	B	A	A	A	A	16
3	A	B	A	A	A	19
4	A	A	B	A	A	22
5	A	A	A	B	A	54
6	A	A	A	A	B	64
7	A	A	A	B	B	29
8	A	A	B	B	B	10
9	A	A	B	A	B	8
10	A	A	B	B	A	7
11	A	B	A	A	B	4
12	A	B	A	B	A	2
13	B	A	B	A	A	5
14	A	B	B	A	A	10
15	A	B	B	A	B	2
16	B	A	A	A	B	4
17	A	A	A	B	C	2
18	A	A	B	A	C	3
19	A	B	A	A	C	6
20	A	B	A	C	A	1
21	B	A	A	A	C	3
22	B	A	A	C	A	1
23	A	A	C	B	B	3
24	A	B	A	B	C	2
25	A	B	A	C	B	2
26	A	B	B	A	C	1
27	A	B	B	C	A	1
28	A	A	B	C	D	1
					Sum:	895

[a] Hu, human; Ch, chimpanzee; Go, gorilla; Or, orangutan; Gi, gibbon. A, B, C, and D are different nucleotides. Data from Brown et al. (1982)

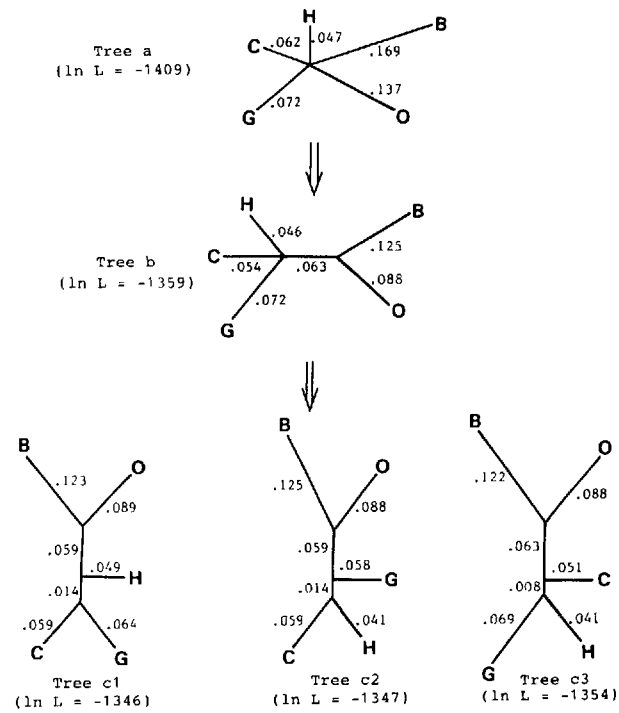

Fig. 8. Three steps to obtain the maximum likelihood tree (c1) for Brown et al.'s (1982) data. H, humans; C, chimpanzees, G, gorillas; O, orangutans; B, gibbons. Branch lengths are not proportional to evolutionary distances.

Table 4. Comparison of three tree-making methods on the probability of obtaining the correct tree of Fig. 9A

(a) One-parameter model					(b) Two-parameter model				
Pattern no.	MP	OD	ML	Obs. no.	Pattern no.	MP	OD	ML	Obs. no.
1	T	T	T	58	1	T	T	T	56
2	T	T	F	9	2	T	T	F	1
3	T	F	T	2	3	T	F	T	1
4	F	T	T	3	4	F	T	T	0
5	T	F	F	1	5	T	F	F	0
6	F	T	F	6	6	F	T	F	4
7	F	F	T	3	7	F	F	T	4
8	F	F	F	18	8	F	F	F	34
% true	70	76	66		% true	58	61	61	

MP, the maximum parsimony method; OD, distance method using the nucleotide difference; ML, the maximum likelihood method; T, the true tree is obtained; F, a false tree is obtained

those for the Jukes-Cantor distance (see also Saitou and Nei 1987). Following Saitou and Nei (1986), both the one-parameter model (Jukes and Cantor 1969) and the two-parameter model (Kimura 1980) were considered. For the latter model, transitions were assumed to occur 10 times more often than transversions. The percentages of obtaining the true tree for the maximum parsimony (MP) method and the distance method (OD) are comparable to the result of Saitou and Nei (1986). Compared to these two methods, the ML method is not so efficient in obtaining the correct topology when the one-parameter model is used. This difference in efficiency is mainly attributable to pattern 2, in which only the ML method failed to reconstruct the correct topology [see Table 4(a)]. On the other hand, all three methods give similar results when the two-parameter model is used.

From the results shown in Table 4, it is possible to consider the relationship between the three tree-making methods in terms of their efficiency in reconstructing correct topologies, as follows. Patterns 2 and 7 suggest the closeness between the maximum parsimony method (MP) and the distance method (OD), because only the ML method failed (pattern 2) or succeeded (pattern 7) in reconstructing the correct topology among three methods. Similarly, patterns 3 and 6 suggest the closeness between the MP method and the ML method, and patterns 4

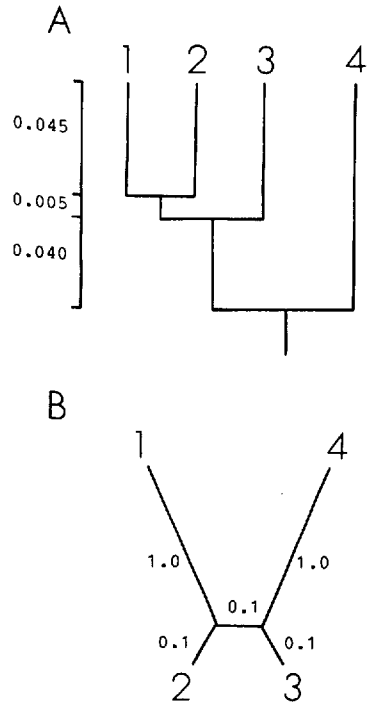

Fig. 9. Two model trees (A and B) used for simulations. Branch lengths are not proportional to evolutionary distances.

and 5 for the distance method (OD) and the ML method. If we compare the frequencies of observing these three types, patterns 4 and 5 were less frequent than others. This suggests that the ML method is more close to the maximum parsimony method than to the distance method. Note also that the ML method chose the same topology as that found by the maximum parsimony method when they were applied to Brown et al.'s (1982) data (see Numerical Examples).

Felsenstein (1978) showed that under a certain condition, the maximum parsimony or compatibility method is positively misleading for finding the correct topology for the case of four OTUs. Figure 9B shows such an example in which branch lengths of pairing OTUs are drastically different ($v_1 = v_4 = 1.0$, $v_2 = v_3 = v_5 = 0.1$) and the constancy of evolutionary rate no longer holds. Assuming the one-parameter model, the probabilities of observing the three phylogenetically informative configurations of nucleotides are computed, and these are 0.0292 for [i, i, j, j], 0.0159 for [i, j, i, j], and 0.0773 for [i, j, j, i]. Here i and j are different nucleotides, and [A, B, C, D] represents the nucleotide configuration for OTUs 1, 2, 3, and 4, respectively. Since the second configuration has the highest probability, we expect to obtain an erroneous tree (OTUs 1 and 4 are clustered) by applying the maximum parsimony or compatibility method.

Assuming the tree of Fig. 9B as the model tree, a computer simulation was done as before, and the maximum parsimony (MP) method and the distance method for the proportion of different nucleotides (OD) and for the Jukes-Cantor distances (ED) are compared with the ML method (see Table 5). Five hundred nucleotides were compared and 100 replications were obtained assuming the one-parameter model. As expected, the MP method always chose the tree in which OTUs 1 and 4 are clustered, instead of the correct tree in which OTUs 1 and 2 are clustered. Interestingly, the distance method also chose the same erroneous tree as in the maximum parsimony method in all cases when the proportions of different nucleotides (OD) were used. When the Jukes-Cantor correction (ED) was made, however, the distance method produced the correct tree with a frequency of 74%. This clear-cut difference between two distances occurred apparently because the former (OD) is metric and behaves like the maximum parsimony score, whereas the expectation of the Jukes-Cantor distance (ED) is proportional to the branch lengths of the model tree.

The ML method chose the correct tree with a frequency of only 43%. The quadrifurcation often gave the highest likelihood. Although there are six cases in which only the ML method chose the true tree (pattern 3), cases in which only the distance method with the corrected distance (ED) outnumbered this (pattern 2). However, the estimates of branch lengths by the ML method were often close to the true values when the true phylogeny was obtained, while the distance method sometimes gave negative estimates of branch lengths (data not shown). Therefore, the ML method may be useful for estimating the branch lengths when the topology is known. Hasegawa and Yano (1984) did a similar study, and they also found that the ML method can obtain the correct topology even when the maximum parsimony method is positively misleading, though they did not consider any distance method.

Discussion

Hixson and Brown (1986) sequenced about 900 nucleotides of a small rRNA gene region for common chimpanzees, pygmy chimpanzees, gorillas, and orangutans. We applied our new algorithm of the ML method for unrooted trees to this data set, including the corresponding human sequence by Anderson et al. (1981). The maximum log-likelihood value for the star tree was −637.9, and the clustering of common and pygmy chimpanzees was chosen at the next step with the log-likelihood of −591.1 out of 10 possibilities. Thus, the search of the ML tree is restricted to the following three topologies. The log-likelihood was −586.0 for tree 1 (chimpanzees and gorillas clustered), −584.0 for tree 2 (humans

Table 5. Comparison of three tree-making methods on the probability of obtaining the correct tree of Fig. 9B

Pattern no.	MP	OD	ML	ED	Obs. no.
1	F	F	T	T	37
2	F	F	F	T	37
3	F	F	T	F	6
4	F	F	F	F	20
% true	0	0	43	74	

ML, the maximum likelihood method; MP, the maximum parsimony method; OD, distance method using the nucleotide difference; ED, distance method using the Jukes-Cantor distance; T, the true tree is obtained; F, a false tree is obtained

and chimpanzees clustered), and −582.9 for tree 3 (humans and gorillas clustered). If we consider the branching pattern of humans, chimpanzees, and gorillas only, trees 1, 2, and 3 correspond to trees c1, c2, and c3 of Fig. 8, respectively. In any case, the final tree becomes as given in Fig. 10A, where humans and gorillas are clustered. In this tree, the root is given by assuming that gibbons diverged first among the five species in hominoid evolution. Branch lengths in parentheses are obtained by averaging estimated (patristic) distances between gibbons and the four remaining species. In this case, a rough constancy of evolutionary rate is assumed.

When we apply other tree-making methods to the same set of data, tree 2 or tree 3 is chosen, depending on the method used. UPGMA and Fitch-Margoliash's method choose tree 2, whereas the transformed distance method and the neighbor-joining method find tree 3 as the best tree. Figure 10B shows the tree reconstructed by the neighbor-joining method. Most of the branch length estimates are quite similar to those of the ML method (Fig. 10A). The location of the root of this tree is obtained in the same way as in Fig. 10A. On the other hand, the distance Wagner method chooses tree 2 when the proportion of different nucleotides (OD) is used, but tree 3 is chosen when the Jukes-Cantor distance (ED) is used. Finally, the maximum parsimony method finds both trees 2 and 3 as equally parsimonious (Hixson and Brown 1986). These inconsistent results indicate that Hixson and Brown's (1986) mitochondrial DNA data are not sufficient to determine the branching order among humans, chimpanzees, and gorillas. This seems to be consistent with the simulation result of Saitou and Nei (1986). Interestingly, however, tree 1, in which chimpanzees and gorillas are clustered, is never chosen by any method, and the log-likelihood of this tree is the smallest among the three trees (see above). Thus, the data of Hixson and Brown (1986) are not favorable for tree 1, and do not support the argument based on a shared one-base deletion between

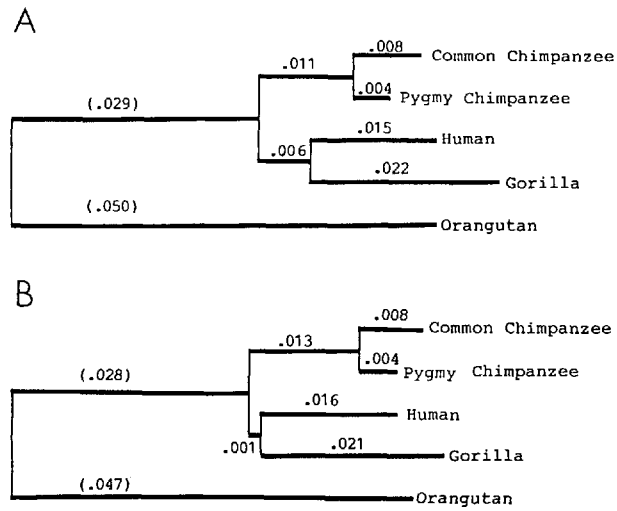

Fig. 10. Two phylogenetic trees reconstructed from Hixson and Brown's (1986) data. Branch lengths are proportional to evolutionary distances. **A** Tree obtained by the maximum likelihood method (new algorithm). **B** Tree obtained by the neighbor-joining method of Saitou and Nei (1987).

chimpanzees and gorillas (Hixson and Brown 1986). Because the nucleotide sequences studied are relatively short, it is difficult to derive a definite conclusion.

In the present study, I examined the statistical properties of the maximum likelihood method and proposed a new algorithm. I also conducted computer simulations to study the relative efficiency of this method compared with other tree-making methods. Although the ML method was shown not to be as efficient as some distance methods, this conclusion is based on the result for unrooted trees for four OTUs. To know the general performance of the ML method, a more extensive study is necessary. It should also be noted that the present study is conducted by using the one- and two-parameter models of nucleotide substitution. The actual pattern of nucleotide substitution is much more complicated than those models and varies from gene to gene (Gojobori et al. 1982b; Li et al. 1984). It is therefore important to study how robust each tree-making method is for various patterns of nucleotide substitution. Further, the rate of nucleotide substitution is assumed to be equal for any nucleotide site in the usual formulation of the maximum likelihood method, whereas the rate varies from site to site in reality. This rate variation may be a serious drawback of the ML method. Clearly, a more detailed study is necessary.

Acknowledgments. I thank Dr. Masatoshi Nei for discussions and helpful comments on this manuscript. This work was supported by grants from the National Science Foundation (USA) and National Institutes of Health (USA) to Dr. Masatoshi Nei,

and by a grant-in-aid (no. 62790190) from the Ministry of Education, Science, and Culture (Japan) to N. Saitou.

References

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465

Brown WM, Prager EM, Wang A, Wilson AC (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution. J Mol Evol 18:225–239

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. Am J Hum Genet 19: 233–257

Chakraborty R (1977) Estimation of time of divergence from phylogenetic studies. Can J Genet Cytol 19:217–223

Eck RV, Dayhoff MO (1966) Atlas of protein sequence and structure 1966. National Biomedical Research Foundation, Silver Spring MD

Faith DP (1985) Distance methods and the approximation of most-parsimonious trees. Syst Zool 34:312–325

Farris JS (1972) Estimating phylogenetic trees from distance matrices. Am Nat 106:645–668

Farris JS (1977) On the phenetic approach to vertebrate classification. In: Hecht MK, Goody PC, Hecht BM (eds) Major patterns in vertebrate evolution. Plenum, New York, pp 823–850

Felsenstein J (1973) Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. Syst Zool 22:240–249

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

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–376

Felsenstein J (1984) The statistical approach to inferring evolutionary trees and what it tells us about parsimony and compatibility. In: Duncan T, Steussy TF (eds) Cladistics: perspectives on the reconstruction of evolutionary history. Columbia University Press, New York, pp 169–191

Fitch WM (1977) On the problem of discovering the most parsimonious tree. Am Nat 111:223–257

Fitch WM (1981) A non-sequential method for constructing trees and hierarchical classifications. J Mol Evol 18:30–37

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155:279–284

Gojobori T, Ishii K, Nei M (1982a) Estimation of average number of nucleotide substitutions when the rate of substitution varies with nucleotide. J Mol Evol 18:414–423

Gojobori T, Li W-H, Graur D (1982b) Patterns of nucleotide substitution in pseudogenes and functional genes. J Mol Evol 18:360–369

Hasegawa M, Yano T (1984) Maximum likelihood method of phylogenetic inference from DNA sequence data. Bull Biometric Soc Jpn 5:1–7

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

Hixson JE, Brown WM (1986) A comparison of the small ribosomal RNA genes from the mitochondrial DNA of the great apes and humans: sequence, structure, evolution, and phylogenetic implications. Mol Biol Evol 3:1–18

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism, vol III. Academic Press, New York, pp 21–132

Kashyap RL, Subas S (1974) Statistical estimation of parameters in a phylogenetic tree using a dynamic model of the substitutional process. J Theor Biol 47:75–101

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

Kimura M (1981) Estimation of evolutionary distances between homologous nucleotide sequences. Proc Natl Acad Sci USA 78:454–458

Klotz LC, Blanken RL (1981) A practical method for calculating evolutionary trees from sequence data. J Theor Biol 91:261–272

Le Quesne WJ (1969) A method of selection of characters in numerical taxonomy. Syst Zool 18:201–205

Li W-H (1981) Simple method for constructing phylogenetic trees from distance matrices. Proc Natl Acad Sci USA 78: 1085–1089

Li W-H (1986) Evolutionary change of restriction cleavage sites and phylogenetic inference. Genetics 113:187–213

Li W-H, Wu C-I, Luo C-C (1984) Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. J Mol Evol 21: 58–71

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nei M, Stephens JC, Saitou N (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

Saitou N, Nei M (1986) The number of nucleotides required to determine the branching order of three species with special reference to the human–chimpanzee–gorilla divergence. J Mol Evol 24:189–204

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sattath S, Tversky A (1977) Additive similarity trees. Psychometrika 42:319–345

Sokal R, Sneath PHP (1963) Principles of numerical taxonomy. WH Freeman, San Francisco

Tajima F, Nei M (1984) Estimation of evolutionary distance between nucleotide sequences. Mol Biol Evol 1:269–285

Takahata N, Kimura M (1981) A model of evolutionary base substitutions and its application with special reference to rapid change of pseudogenes. Genetics 98:641–657

Tateno Y, Nei M, Tajima F (1982) Accuracy of estimated phylogenetic trees from molecular data. I. Distantly related species. J Mol Evol 18:387–404

Received August 21, 1987/Revised December 7, 1987