[36] Maximum Likelihood Methods

By NARUYA SAITOU

General Principle

Application of the maximum likelihood (ML) method to the problem of phylogenetic tree reconstruction was first studied for the case of gene frequency data.[1] Later, an ML algorithm for constructing unrooted phylogenetic trees from nucleotide sequence data was developed by Felsenstein.[2] Recently, Saitou[3] proposed a stepwise tree-searching algorithm for the ML method. This is similar to that of the neighbor-joining method,[4] in which distance matrices are used.

Let us first explain the general principle of the ML method for nucleotide sequence data. Consider the tree in Fig. 1. Let us ignore the root R and consider the tree as an unrooted tree. We first restrict our attention to a specific nucleotide site, and assume that nucleotide N_i was observed at sequence (node) i (i = A, B, C, D, or E). On the other hand, nucleotide N_j at node j (j = X, Y, or Z) is unknown, and it can be one of four nucleotides. Thus, the likelihood (L) for this site becomes

$$L = \sum_{N_Y} \left[g_Y P_{YC} \left(\sum_{N_X} P_{YX} P_{XA} P_{XB} \right) \left(\sum_{N_Z} P_{YZ} P_{ZD} P_{ZE} \right) \right] \qquad (1)$$

where g_Y is the probability that node Y has nucleotide N_Y, $P_{ij} \equiv P(N_i,N_j,d_{ij})$ is the probability of observing nucleotides N_i and N_j in sequences i and j, respectively, and d_{ij} is the expected number of nucleotide substitutions between these two sequences. Summation is for four possible nucleotides, because N_X, N_Y, and N_Z are variables. To obtain $P(N_i,N_j,d_{ij})$, we must specify the pattern of nucleotide substitution. If we use the random substitution model (the one-parameter model),[5]

$$P(N_i,N_j,d_{ij}) = \tfrac{1}{4} + (\tfrac{3}{4}) \exp(-4d_{ij}/3) \qquad (\text{if } N_i = N_j) \qquad (2a)$$

$$P(N_i,N_j,d_{ij}) = \tfrac{1}{4} - (\tfrac{1}{4}) \exp(-4d_{ij}/3) \qquad (\text{if } N_i \neq N_j) \qquad (2b)$$

[1] L. L. Cavalli-Sforza and A. W. F. Edwards, *Am. J. Hum. Genet.* **19**, 233 (1967).
[2] J. Felsenstein, *J. Mol. Evol.* **17**, 368 (1981).
[3] N. Saitou, *J. Mol. Evol.* **27**, 261 (1988).
[4] N. Saitou and M. Nei, *Mol. Biol. Evol.* **4**, 406 (1987).
[5] T. H. Jukes and C. R. Cantor, *in* "Mammalian Protein Metabolism" (H. N. Munro, ed.), p. 21. Academic Press, New York, 1969.

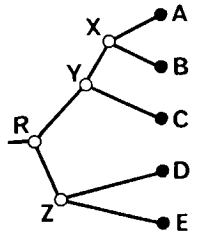

FIG. 1. Phylogenetic tree for five nucleotide sequences.

If we assume the two-parameter model,[6] in which transitions and transversions can occur at different rates, the expressions that correspond to Eqs. (2a) and (2b) are somewhat complicated [see Eqs. (9a)–(9c)]. It is also true when other kinds of data such as amino acid sequences[7] or restriction site data[8] are considered. However, the essential nature of the likelihood function remains the same.

In any case, the likelihood for each nucleotide site defined by Eq. (1) is then multiplied for all sites and computed for many combinations of branch lengths for a given tree topology, and the combination that shows the highest likelihood is chosen as the ML solution. It has been proved that there exists a single ML point in the possible parameter range for a given tree topology when the one-parameter model of nucleotide substitution was assumed.[9]

For the case of a rooted tree, it is necessary to assume a constant rate of nucleotide substitution. For example, the following constraints should be invoked for the tree in Fig. 1: $d_{XA} = d_{XB}$, $d_{ZD} = d_{ZE}$, $d_{YC} = d_{YX} + d_{XA}$, and $d_{RY} + d_{YC} = d_{RZ} + d_{ZD}$. Because of this, the number of parameters to be estimated for n sequences is $n - 1$ for rooted trees, compared to $2n - 3$ parameters (branch lengths) for unrooted trees.

Each nucleotide site is considered separately in Felsenstein's method.[2] When each site is assumed to evolve at the same evolutionary rate, however, a more essential unit of comparison for the ML method is the "nucleotide configuration." A nucleotide configuration is the distribution pattern of nucleotides for a given set of sequences (see Tables I and II for examples). When the one-parameter model is assumed, the possible number of configurations for n sequences is $(4^{n-1} + 3 \times 2^{n-1} + 2)/6$.[10] There

[6] M. Kimura, *J. Mol. Evol.* **16,** 111 (1980).
[7] R. L. Kashyap and S. Subas, *J. Theor. Biol.* **47,** 75 (1974).
[8] P. E. Smouse and W.-H. Li, *Evolution* **41,** 1162 (1987).
[9] K. Fukami and Y. Tateno, *J. Mol. Evol.* **28,** 460 (1989).
[10] N. Saitou and M. Nei, *J. Mol. Evol.* **24,** 189 (1986).

TABLE I
NUCLEOTIDE CONFIGURATIONS FOR THREE SEQUENCES

Configuration[a]				Observed no.	
No.	A	B	C	Case a	Case b
1	i	i	i	789	40
2	i	i	j	98	20
3	i	j	i	59	15
4	j	i	i	50	5
5	i	j	k	4	20
				1,000	100

[a] A, B, and C are different sequences, and i, j, and k are nucleotides that are different from each other.

TABLE II
NUCLEOTIDE CONFIGURATIONS FOR FOUR SEQUENCES

Configuration[a]					
No.	A	B	C	D	Observed no.
1	i	i	i	i	196
2	i	i	i	j	67
3	i	i	j	i	40
4	i	j	i	i	77
5	j	i	i	i	22
6	i	i	j	j	12
7	i	j	i	j	5
8	i	j	j	i	8
9	i	i	j	k	11
10	i	j	i	k	21
11	j	i	i	k	10
12	j	k	i	i	12
13	j	i	k	i	3
14	i	j	k	i	15
15	i	j	k	l	1
					500

[a] A, B, C, and D are sequences of Fig. 4, and i, j, k, and l are different nucleotides.

are 5, 15, and 51 configurations for three, four, and five sequences, respectively.

If we consider the nucleotide configuration, the general formula for the likelihood becomes

$$L = \prod_{i=1}^{c} U_i^{m_i} \times \frac{m!}{m_1!m_2! \ldots m_c!} \tag{3}$$

where c is the number of possible nucleotide configurations, U_1 the probability of obtaining the ith configuration, m_i the observed number of the ith configuration, and m the sum of m_i values, or the total number of nucleotides compared. Because the number of configurations depends on the number of sequences compared, we consider the cases of two, three, four, and five sequences separately. The case of unrooted trees is first considered.

Two Sequences

When we have only two sequences, there are only one unrooted and one rooted trees. Thus the ML solution for the rooted tree becomes identical to that for the unrooted tree.

Let us consider the one-parameter model of nucleotide substitution. There are only two configurations for this case: two nucleotides are the same (configuration 1) or different each other (configuration 2). Therefore, Eq. (3) becomes

$$L = U_1^{m_1} \times U_2^{m_2} \times \frac{(m_1 + m_2)!}{m_1!m_2!} \tag{4}$$

where $U_1 = 1/4 + (3/4)\exp(-4d_{ij}/3)$ [see Eq. (2a)], and $U_2 = 3/4 - (3/4)\exp(-4d_{ij}/3)$ [see Eq. (2b)]. First, we take the logarithm,

$$\log L = m_1 \log[(1 + 3x)/4] + m_2 \log[(3 - 3x)/4] + \text{constant} \tag{5}$$

where $x = \exp(-4d_{ij}/3)$. To obtain the maximum likelihood estimate of d_{ij}, we differentiate $\log L$ with respect to x, and equate it to zero. This gives

$$\frac{-4m_1x}{1 + 3x} + \frac{4m_2x}{3 - 3x} = 0 \tag{6}$$

Solving Eq. (6) (note that $x \neq 0$ while d_{ij} is finite),

$$\hat{x} = 1 - 4m_2/3(m_1 + m_2) \tag{7}$$

Hence,

$$d_{ij} = -(3/4)\log(1 - 4\pi_{ij}/3) \tag{8}$$

where $\pi_{ij} = m_2/(m_1 + m_2)$. Equation (8) is identical to the famous formula of Jukes and Cantor.[5]

When the two-parameter model[6] is assumed, we have to distinguish transitional and transversional differences. Thus, the number of configurations becomes three. Probabilities U_i ($i = 1, 2, 3$) are given by

$$U_1 = 1/4 + p^2/4 + pq/2 \qquad \text{(no difference)} \tag{9a}$$

$$U_2 = 1/4 + p^2/4 - pq/2 \qquad \text{(transitional difference)} \tag{9b}$$

$$U_3 = 1/2 - p^2/2 \qquad \text{(transversional difference)} \tag{9c}$$

where $p = \exp(-2\beta)$ and $q = \exp(-2\alpha)$. α and β are the expected numbers of transitions and transversions, respectively. A likelihood function can be obtained using these U_i values, and we equate the differentiations of the log likelihood with p and q to be zero, as in the case of the one-parameter model. We solve the two equations,

$$\hat{p} = (1 - 2m_3/m)^{1/2} \tag{10a}$$

$$\hat{q} = [1 - (2m_2 + m_3)/m]/\hat{p} \tag{10b}$$

where m_i is the observed number of the ith configuration and $m = m_1 + m_2 + m_3$. Thus, the total number of nucleotide substitutions $d \equiv \alpha + 2\beta$ is estimated by

$$\begin{aligned} \hat{d} &= -(1/2)\log(\hat{p}^2\hat{q}) \\ &= -(1/2)\log[(1 - 2P - Q)(1 - 2Q)^{1/2}] \end{aligned} \tag{11}$$

where $P = m_2/m$ and $Q = m_3/m$. Equation (11) is equivalent to Kimura's estimation formula of evolutionary distance.[6]

It is hypothesized that a similar correspondence between the ML estimate and the distance formula above can be applied to other types of nucleotide substitution models (see this volume, [33] for other models). If this conjecture is true, the ML method is identical with the distance method for the case of two sequences.

Three Sequences

Unrooted Trees

There is only one unrooted tree for three sequences (Fig. 2a). Let us assume the one-parameter model of nucleotide substitution. There are five nucleotide configurations (Table I). Probabilities U_i for a given set of three distances (d_{XA}, d_{XB}, and d_{XC}) are given by

$$U_1 = 4h(i,i,i), \tag{12a}$$

$$U_2 = 12h(i,i,j) \qquad (i \neq j) \tag{12b}$$

$$U_3 = 12h(i,j,i) \qquad (i \neq j) \tag{12c}$$

$$U_4 = 12h(j,i,i) \qquad (i \neq j) \tag{12d}$$

$$U_5 = 24h(i,j,k) \qquad (i \neq j \neq k) \tag{12e}$$

where i, j, and k are any one of the four nucleotides and

$$h(i,j,k) = \sum_x g_x P_{xi}(d_{XA}) P_{xj}(d_{XB}) P_{xk}(d_{XC}) \tag{13}$$

where g_x is the probability of observing nucleotide x at node X and $P_{xi}(d_{XA}) = P(x,i,d_{XA})$ [see Eqs. (2a) and (2b)]. By substituting these U_i values into Eq. (3), the likelihood function is obtained.

An analytical solution for the ML value is not easily found, but the ML value can be obtained numerically by changing three distances (d_{XA}, d_{XB}, and d_{XC}). In this case, however, the ML estimates of these distances are no longer the same as those obtained by the distance method,[11] in which case the branch length estimates are given by:

$$\hat{d}_{AX} = (d_{AB} + d_{AC} - d_{BC})/2 \tag{14a}$$

$$\hat{d}_{BX} = (d_{AB} + d_{BC} - d_{AC})/2 \tag{14b}$$

$$\hat{d}_{CX} = (d_{AC} + d_{BC} - d_{AB})/2 \tag{14c}$$

where d_{AB}, d_{AC}, and d_{BC} are given using Eq. (8), which is the ML estimate of each distance value. In applying Eq. (8) the proportion of nucleotide differences (π_{ij}) between sequences i and j is estimated by

$$\pi_{AB} = (m_3 + m_4 + m_5)/m \tag{15a}$$

$$\pi_{AC} = (m_2 + m_4 + m_5)/m \tag{15b}$$

$$\pi_{BC} = (m_2 + m_3 + m_5)/m \tag{15c}$$

It is clear from Eqs. (15a)–(15c) that different sets of m_i values ($1 \leq i \leq 5$) can give the same set of π_{AB}, π_{AC}, and π_{BC} values. Thus, the estimates of d_{XA}, d_{XB}, and d_{XC} by the distance method are not the ML estimates. However, it has been shown that the ML estimates of branch lengths are quite similar to those obtained by the distance method.[12]

Rooted Trees

There are three rooted trees for three sequences, and we have two parameters (d_1 and d_2) to be estimated (see Fig. 2b). Let us designate the tree in Fig. 2b as "tree 1," in which sequences A and B are closer to each other than to C. As for the other two possible trees, sequences A and C are closer in tree 2, and B and C are closer in tree 3.

[11] W. M. Fitch and E. Margoliash, *Science* **155**, 279 (1967).

[12] N. Saitou, Ph.D. dissertation, University of Texas at Houston, 1986.

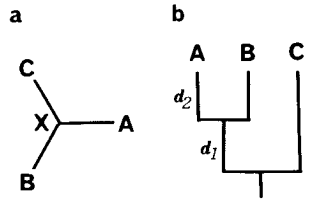

FIG. 2. (a) Unrooted tree and (b) rooted tree for three sequences.

Assuming the one-parameter model, expressions for U_i for tree 1 become

$$U_1 = (A + Ba^2)/16 \tag{16a}$$
$$U_2 = (3A - Ba^2)/16 \tag{16b}$$
$$U_3 = (C + Da^2)/16 \tag{16c}$$
$$U_4 = U_3 \tag{16d}$$
$$U_5 = 2(C - Da^2)/16 \tag{16e}$$

where $A = 1 + 3b^2$, $B = 6(1 + b)b^2$, $C = 3(1 - b^2$, $D = 6(1 - b)b^2$, and $a = \exp(-4d_1/3)$ and $b = \exp(-4d_2/3)$.[3] [Equation (16e) corresponds to Eq. (4e) of Saitou,[3] which was incorrect.] From these probabilities, one can evaluate the likelihood, $L(1)$, of tree 1 under specific d_1 and d_2 values. Computation of $L(2)$ and $L(3)$ for trees 2 and 3, respectively, is done in a similar manner.

Saitou[3] derived the conditions for obtaining tree 1 as the ML estimate, or for obtaining the relationship $L(1) > L(2)$ and $L(1) > L(3)$, i.e., $m_2 > m_3$ and $m_2 > m_4$, where m_i is the observed number of the ith nucleotide configuration (see Table I). When this condition is satisfied, d_{AB} becomes the smallest among d_{AB}, d_{AC}, and d_{BC}. Thus, tree 1 is chosen if we use UPGMA (unweighted pair-group method),[13] in which the pair of sequences with the smallest distance is first clustered. This means that the topology of the UPGMA tree is always identical with that of the maximum likelihood tree, though the estimates of branch lengths (d_1 and d_2) may be different to some extent.

Figure 3 shows two examples of likelihood surfaces for artificial data (Cases a and b of Table I).[3] In Case a, the above condition ($m_2 > m_3$ and $m_2 > m_4$) is satisfied and tree 1 has the ML value at $d_1 = 0.0259$, whereas the trifurcating tree ($d_1 = 0$) gives the ML value for the other two trees (Fig.

[13] R. Sokal and P. H. A. Sneath, "Principles of Numerical Taxonomy." Freeman, San Francisco, California, 1963.

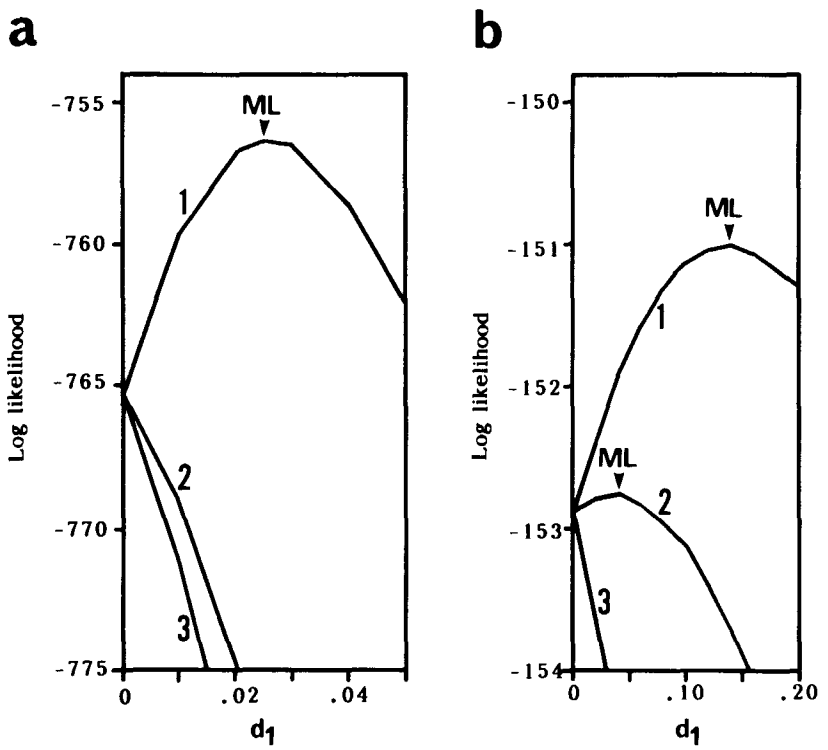

FIG. 3. Two examples of likelihood surfaces for three possible trees. (a) and (b) correspond to Cases a and b in Table I. See text for details.

3a). The likelihood surfaces represent the ML values for given d_1 values (an ML value was computed by varying d_2 for a given d_1 value).

The above condition is also satisfied in Case b, and tree 1 gives the highest ML value. However, tree 2 also gives an ML value higher than that for the trifurcating tree (Fig. 3b). In this case the following relationship is satisfied:

$$d_{AC} < (d_{AB} + d_{BC})/2 \tag{17}$$

Thus, if sequences A and C are first clustered (tree 2), a positive estimate of d_1 is obtained. This situation is different from that for Case a, in which a negative estimate of d_1 is obtained if tree 2 is considered. It seems that a positive estimate of d_1 corresponds to the ML value higher than that for the trifurcating tree.[3] For tree 3, however, d_1 becomes negative, and the log-likelihood value decreases as the value of d_1 increases (Fig. 3b).

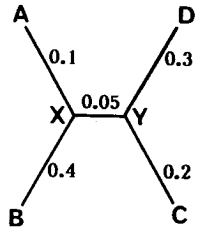

FIG. 4. Unrooted tree for four sequences. Numbers are lengths for each branch.

Four Sequences

Unrooted Trees

There are three unrooted trees for four sequences, and we have to consider 15 nucleotide configurations (Table II). The likelihood function (not shown) can be obtained in a manner similar to the case of three sequences.

Let us consider a numerical example. This is a result based on a simulated data set (see Table II), assuming the tree of Fig. 4.[3] The tree in which sequences A and B are clustered (i.e., the true tree as in Fig. 4) had the highest ML value (log likelihood is -1004.2) with $\hat{d}_{XY} = 0.026$. The estimates of other branch lengths were $\hat{d}_{XA} = 0.110$, $\hat{d}_{XB} = 0.361$, $\hat{d}_{YC} = 0.188$, and $\hat{d}_{YD} = 0.303$. On the other hand, the same tree was obtained when we applied the neighbor-joining method for a distance matrix. Branch length estimates were $\hat{d}_{XY} = 0.033$, $\hat{d}_{XA} = 0.101$, $\hat{d}_{XB} = 0.366$, $\hat{d}_{XC} = 0.179$, and $\hat{d}_{XD} = 0.316$, which were close to the estimates obtained by the ML method. The corresponding log-likelihood value for these estimates was -1004.5, slightly lower than that of the ML estimate.

The maximum likelihood value for the other two unrooted trees was obtained for the case with no interior branch (a quadrifurcating tree). This is analogous to the trifurcation of the rooted tree for three sequences. However, other trees may have ML values with positive estimates of the interior branch.

Rooted Trees

The number of possible rooted trees for four sequences is 15, and 3 of them (trees 3a, 3b, and 3c) are shown in Fig. 5. U_i of Eq. (3) is computed in a manner similar to the case of three sequences, and the likelihood function can be obtained using these U_i values.[3]

Although the ML method and UPGMA may no longer give the same tree on every occasion, it has been shown that the tree estimated by the ML

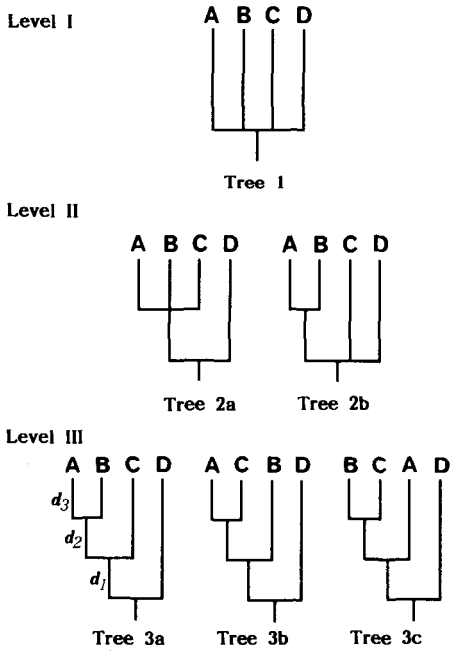

FIG. 5. Three levels of the topological relationships for rooted trees for four sequences.

method is similar to that obtained by UPGMA.[3] In this connection, it should be noted that UPGMA gives least-squares estimates of branch lengths for the tree obtained.[14] That is, UPGMA minimizes the quantity, $\Sigma(d_{ij} - 2\lambda t_{ij})^2$, where λ is the rate of nucleotide substitution and t_{ij} is the time since divergence between sequences i and j.

Let us explain Saitou's algorithm[3] using Fig. 5. This algorithm is reminiscent of the neighbor-joining method. First the ML value for tree 1 (Level I) is computed. There is no information on the clustering of sequences with tree 1, whereas there are trifurcating and bifurcating points in trees at the next step (Level II). The ML values for all 10 trees at Level II are computed and are compared with each other. We choose the tree with the highest ML value among the 10. Suppose it is tree 2a. We then compute the ML values for trees 3a, 3b, and 3c, which are produced when the trifurcating point of tree 2a is resolved. In any case, the tree with the highest ML value is chosen as the final tree. In this way, we may be able to find the tree with the highest ML value. When the number of sequences is

[14] R. Chakraborty, *Can. J. Genet. Cytol.* **19**, 217 (1977).

large, this algorithm requires much less computational time compared with tree-by-tree examination.

The two trees (2a and 2b) at Level II are related to tree 3a at Level III, and they may be used as null trees to test the significance of d_1 and d_2 of tree 3a. If d_1 is not significantly greater than zero, we assume that tree 2a is the correct one and proceed to test the significance of d_2 with tree 1 as the null tree. In general, there are $n-1$ levels for trees with n sequences, and we can successively test the significance of all interior branches of a tree.

Five Sequences

There are 15 unrooted trees and 105 rooted trees for five sequences, and the number of possible nucleotide configurations under the one-parameter model of nucleotide substitution is now 51. Construction of the likelihood function is straightforward as before. We consider only unrooted trees.

Algorithm for Finding ML Tree

Let us explain Saitou's algorithm[3] for the ML method for unrooted trees, with a numerical example of five nucleotide sequences of higher primates (human, common chimpanzee, pygmy chimpanzee, gorilla, and orangutan).[15] As in the case of rooted trees, we first compute the ML estimate for the Level I tree of Fig. 6. The initial value (d_{xi}) for the length of the branch between node X and sequence i may be computed from the distance matrix, applying the neighbor-joining method as follows:[3]

$$d_{xi} = \sum_{j=1}^{5} d_{ij} - \frac{3}{4} \sum_{j<k} d_{jk} \tag{18}$$

The maximum log-likelihood value for the Level I tree was -638 for the example data. Ten different trees (Level II trees) are then considered, and the ML value for each tree is computed. These trees have one trifurcation and one bifurcation. The clustering of common and pygmy chimpanzees gave the highest log-likelihood value of -591. From each tree at Level II, three trees are produced if the trifurcation is resolved (Level III). The ML estimates of six branch lengths for the case of Level II are used as the initial values for the computation of the ML estimation for a tree at Level III (see Fig. 6). In this case, the seven-dimensional likelihood surface is numerically examined for three possible trees. The log-likelihood values were -586 for tree 1 (chimpanzee and gorilla clustered), -584 for tree 2

[15] J. Hixson and W. M. Brown, *Mol. Biol. Evol.* **3**, 1 (1986).

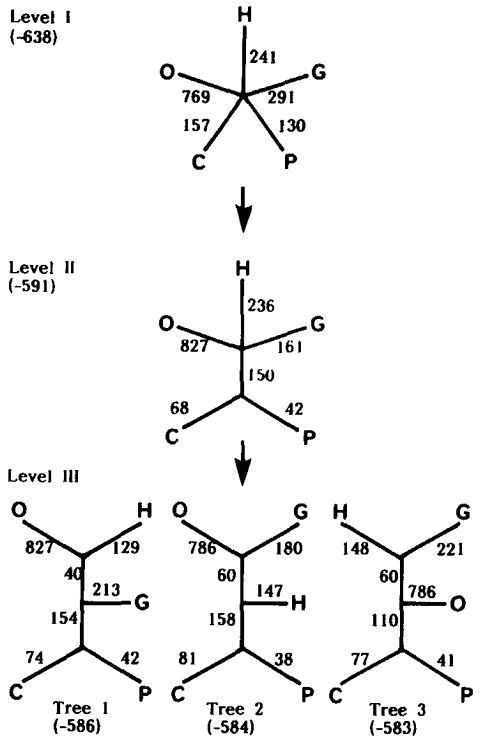

FIG. 6. Three steps to obtain the maximum likelihood tree for the data of Hixson and Brown.[16] H, Human; C, common chimpanzee; P, pygmy chimpanzee; G, gorilla; O, orangutan. Estimated branch lengths for each tree should be divided by 10,000. Figures in parentheses are the maximum log-likelihood values.

(human and chimpanzee clustered), and −583 for tree 3 (human and gorilla clustered). Thus, tree 3 was found to be the ML tree under this algorithm. This new algorithm can be extended to any number of sequences, as in the case for rooted trees.

Comparison with Other Methods

When we applied other tree-making methods to the same data set, the choice of the tree depended on the method used.[3] UPGMA chose tree 1; the neighbor-joining method found tree 3 as the best tree, and most of the branch lengths estimated by this method were quite similar to those of the ML method. On the other hand, the distance Wagner method[16] chose tree 2 when the proportion of different nucleotides was used as distance.

[16] J. Farris, *Am. Nat.* **106**, 645 (1972).

TABLE III
RESULTS OF FOUR METHODS FOR PRIMATE SEQUENCE DATA[a]

Method	Tree 1	Tree 2	Tree 3	Tree 4
Fitch–Margoliash	0	+0.60	+0.47	+25.33
Minimum evolution	+1.10	+0.35	0	+6.59
Maximum parsimony	+1	0	0	+8
Maximum likelihood	−3.97	−2.98	0	−33.86

[a] Values for the Fitch–Margoliash and minimum evolution methods are percent standard deviation and sum of branch lengths × 1000, respectively. Values for the maximum parsimony method are the required number of nucleotide substitutions, and those for the maximum likelihood method are the log likelihood. Values of the best tree are set to zero, and the other values represent differences from the best tree. From Saitou and Imanishi.[17]

Phylogenetic trees are constructed step by step for methods such as UPGMA, the neighbor-joining method, and the distance Wagner method, and finally a single tree is obtained. This type of method can be called a "stepwise" clustering method.[17] In contrast, all possible trees (or a limited number of plausible trees) are compared under a certain criterion for Fitch–Margoliash method,[11] the maximum parsimony method,[18] the minimum evolution method,[17] and the standard ML method.[2] Results of these methods applied for the same data set as above are presented in Table III.[17] Trees 1–3 in Table III correspond to those of Fig. 6, whereas pygmy chimpanzee and human as well as gorilla and orangutan are clustered in tree 4.

The program DNAML of PHYLIP Version 3.1, developed by J. Felsenstein, was used for the ML method. The transition/transversion ratio was set to be 5.0, and observed frequencies of nucleotides were used. Trees 1–4 were examined using the "user tree" option. Because a different model of nucleotide substitution was used in this case, the log-likelihood values for trees 1–3 are different from those obtained using Saitou's algorithm (Fig. 6). However, the order of the maximum log-likelihood values among the three trees is the same as that obtained above, and tree 3 was chosen. The ML value for tree 4 was considerably lower than those for the other three trees. This is apparently because the highly probable clustering of common and pygmy chimpanzees was not realized in this tree.

As for the other methods, trees 1 and 3 were chosen by the method of

[17] N. Saitou and T. Imanishi, *Mol. Biol. Evol.* **6**, 514 (1989).
[18] W. M. Fitch, *Am. Natur.* **111**, 223 (1977).

Fitch–Margoliash and the minimum evolution method, respectively. Interestingly, the rank of these trees in terms of the minimum evolution method was the same as that of the ML method (see Table III). Finally, the maximum parsimony method found both trees 2 and 3 as equally parsimonious. These inconsistent results indicate that the nucleotide sequence data used are not sufficient to determine the branching order among human, chimpanzee, and gorilla. This seems to be consistent with the simulation result of Saitou and Nei.[10]

Discussion

The original formulation of the ML method[1] included the probability of tree topology, which is given by assuming a Yule process. Felsenstein[2] took a different approach, in which the ML value for each tree is computed and the tree with the highest ML value is chosen. However, the likelihood function to be used varies from tree to tree, so the ML values for different trees are conditional and cannot be compared in the standard way.[3,19] It has been claimed[20] that this problem can be avoided by applying information theory. Yet there still remains a paradox.

Let us consider Fig. 3b. In this case both trees 1 and 2 had higher ML values than that for the trifurcating tree where d_1 is necessarily zero. Because this trifurcating tree is a submodel of either tree 1 or 2, we may use the likelihood ratio test for the significance of these two trees. As Felsenstein[21] indicated, however, it is possible that both trees 1 and 2 are significantly better than the trifurcating tree. Then which tree should we choose? This paradox seems to apply in trees 1–3 of Fig. 6, since the log-likelihood values for all these trees are considerably larger than that of Level II tree. Therefore, we should be cautious in applying the ML method for the problem of phylogenetic tree estimation.

Nevertheless, the ML method may still be useful for practical purposes. For example, it has been shown by computer simulations[3,22] that the ML method can find the correct tree with an appreciable proportion even when the maximum parsimony method is positively misleading under certain conditions for four sequences.[23] Saitou and Imanishi[17] compared the relative efficiency of the five tree-making methods (maximum parsimony, maximum likelihood, Fitch–Margoliash, minimum evolution, and neigh-

[19] M. Nei, "Molecular Evolutionary Genetics." Columbia Univ. Press, New York, 1987.
[20] T. Kishino and M. Hasegawa, *J. Mol. Evol.* **29**, 170 (1989).
[21] J. Felsenstein, *in* "Statistical Analysis of DNA Sequence Data" (B. S. Weir, ed.), p. 133. Dekker, New York, 1983.
[22] M. Hasegawa and T. Yano, *Bull. Biomed. Soc. Jpn.* **5**, 1 (1984).
[23] J. Felsenstein, *Syst. Zool.* **27**, 401 (1978).

bor-joining methods). They used a computer simulation under the model tree for six sequences, and showed that the neighbor-joining method, the minimum evolution method, and the ML method performed better than Fitch–Margoliash and maximum parsimony methods.

From a practical point of view, however, a distance matrix method, such as the neighbor-joining method, seems to be the first choice for determining tree topology. The maximum likelihood method is time consuming, and should better be used after a certain number of prospective trees are chosen by some distance methods. For a review of distance matrix methods for phylogenetic tree construction, readers may refer to Nei[19] and Saitou.[24]

[24] N. Saitou, *in* "Handbook of Statistics, Volume 8: Statistical Methods for Biological and Medical Sciences" (C. R. Rao and R. Chakraborty, eds.). North-Holland, New York, in press.