J Mol Evol (1983) 19:153–170

Journal of Molecular Evolution

Accuracy of Estimated Phylogenetic Trees from Molecular Data

II. Gene Frequency Data

Masatoshi Nei, Fumio Tajima, and Yoshio Tateno*

Center for Demographic and Population Genetics, The University of Texas at Houston, Houston, Texas 77025

Summary. The accuracies and efficiencies of three different methods of making phylogenetic trees from gene frequency data were examined by using computer simulation. The methods examined are UPGMA, Farris' (1972) method, and Tateno et al.'s (1982) modified Farris method. In the computer simulation eight species (or populations) were assumed to evolve according to a given model tree, and the evolutionary changes of allele frequencies were followed by using the infinite-allele model. At the end of the simulated evolution five genetic distance measures (Nei's standard and minimum distances, Rogers' distance, Cavalli-Sforza's f_θ, and the modified Cavalli-Sforza distance) were computed for all pairs of species, and the distance matrix obtained for each distance measure was used for reconstructing a phylogenetic tree. The phylogenetic tree obtained was then compared with the model tree. The results obtained indicate that in all tree-making methods examined the accuracies of both the topology and branch lengths of a reconstructed tree (rooted tree) are very low when the number of loci used is less than 20 but gradually increase with increasing number of loci. When the expected number of gene substitutions (M) for the shortest branch is 0.1 or more per locus and 30 or more loci are used, the topological error as measured by the distortion index (d_T) is not great, but the probability of obtaining the correct topology (P) is less than 0.5 even with 60 loci. When M is as small as 0.004, P is substantially lower. In obtaining a good topology (small d_T and high P) UPGMA and the modified Farris method generally show a better performance than the Farris method. The poor performance of the Farris method is observed even when Rogers' distance which obeys the triangle inequality is used. The main reason for this seems to be that the Farris method often gives overestimates of branch lengths. For estimating the expected branch lengths of the true tree UPGMA shows the best performance. For this purpose Nei's standard distance gives a better result than the others because of its linear relationship with the number of gene substitutions. Rogers' or Cavalli-Sforza's distance gives a phylogenetic tree in which the parts near the root are condensed and the other parts are elongated. It is recommended that more than 30 loci, including both polymorphic and monomorphic loci, be used for making phylogenetic trees. The conclusions from this study seem to apply also to data on nucleotide differences obtained by the restriction enzyme techniques.

Key words: UPGMA – Farris' method – Modified Farris method – Genetic distance – Topological errors – Errors in branch length – Triangle inequality.

Introduction

In the previous paper (Tateno et al. 1982) we studied the accuracies of four different methods of constructing phylogenetic trees for molecular data by using computer simulation. The strategy used was to simulate the evolutionary changes of nucleotide sequences for a given number of species and compare the phylogenetic tree reconstructed from simulated sequence data with the hypothetical model tree following which the species evolved. In this study we were primarily concerned with long-term evolution, so that the polymorphism within species was ignored. When a phylogenetic tree for closely related

Present address: Institute of Physical and Chemical Research, Rikagaku Kenkyusho, Hirosawa, Wako-shi, Saitama, 351, Japan

Offprint requests to: Masatoshi Nei

species or populations is constructed, however, the effect of polymorphism cannot be neglected. In this case we must consider all polymorphic alleles and their frequency changes in populations and measure the genetic distance between populations statistically. The purpose of this paper is to conduct this type of study and compare the accuracies and efficiencies of different tree-making methods for allele frequency data.

A somewhat similar study was previously conducted by Kidd and Cavalli-Sforza (1971). These authors assumed that the evolutionary change of their "character" follows the Brownian motion, so that each character showed the normal distribution in all generations. However, it is known that the allele frequency change is typically non-Brownian, though it *can* be approximated by the Brownian motion under special circumstances in terms of arcsine or normid transformation (Cavalli-Sforza and Edwards 1967; Cavalli-Sforza and Piazza 1975). Furthermore, they did not consider new mutation or extinction of alleles which inevitably occur in the evolutionary process. Therefore, their theories are not directly applicable to actual allele frequency data. In the following we shall consider a more realistic model of the genetic change of populations by using Kimura and Crow's (1964) infinite-allele model of mutation.

In the previous paper we examined four different tree-making methods, i.e., UPGMA (Sneath and Sokal 1973), Fitch and Margoliash's (1967) method, Farris' (1972) method, and Tateno et al.'s (1982) modified Farris method. In the present paper we shall exclude Fitch and Margoliash's method, because it requires a large amount of computer time and our previous study has shown that it is not as efficient as the modified Farris method.

Some of the preliminary results of this study have been published by Tateno (1982). He used Nei's genetic distance and his own distortion index for comparing the topological errors of the trees reconstructed by the above three methods. His results showed that the distortion indices of the three methods are more or less the same. In this paper we present the results of our comprehensive study on the relative merits of the three tree-making methods by using Robinson and Foulds' (1981) distortion index and five distance measures. We shall show that UPGMA with Nei's standard distance generally gives a better tree than the other two methods.

Models and Methods of Computer Simulations

1) Genetic Distance. One of the most frequently used measures of genetic distance for molecular data is Nei's (1972) standard genetic distance (D). This measure is intended to estimate the average number of codon (gene) substitutions per locus. In the study of genetic relationship among closely related organisms allelic differences are usually studied by electrophoresis. In this case the genetic distance obtained is expected to be considerably lower than the actual number of codon substitutions, since electrophoresis does not detect all codon differences between alleles. However, D is expected to be proportional to the number of codon substitutions, unless the distance is very large (Nei 1978a).

Consider two populations, X and Y, and let x_{ij} and y_{ij} be the frequencies of the i-th allele at the j-th locus in populations X and Y, respectively. D is then defined as

$$D = -\log_e (J_{XY}/\sqrt{J_X J_Y}) \quad , \tag{1}$$

where

$$J_X = \sum_j^r \sum_i^{m_j} x_{ij}^2/r \quad , \quad J_Y = \sum_j^r \sum_i^{m_j} y_{ij}^2/r \quad ,$$

and

$$J_{XY} = \sum_j^r \sum_i^{m_j} x_{ij}y_{ij}/r \quad ,$$

in which m_j is the number of alleles at the j-th locus and r is the number of loci studied. Using the infinite-allele model of mutation (Kimura and Crow 1964), Nei (1972) has shown that the expectation of D when a large number of loci are used is given by

$$E(D) = 2vt \quad , \tag{2}$$

where v is the mutation rate or the rate of gene substitution per locus per generation and t is the time since divergence between populations X and Y. The sampling variance (Nei and Roychoudhury 1974; Nei 1978b) and drift variance (Li and Nei 1975) of D have also been worked out.

Recently Farris (1981) claimed that D is not appropriate for making a phylogenetic tree, because it is not a metric and does not obey the triangle inequality. It is therefore interesting to see whether this measure gives a poor performance in tree-making compared with metric distances such as D_R in (5). It should be noted that the nonmetricity of D occurs because of the stochastic nature of gene substitution. Theoretically, as the number of loci used increases, D gradually attains metricity. In other words, it is an asymptotic metric.

A measure related to D is Nei's (1973) minimum genetic distance (D_m). It is defined as

$$D_m = \frac{1}{2r} \sum_{j=1}^r \sum_{i=1}^{m_j} (x_{ij} - y_{ij})^2 \quad . \tag{3}$$

The expectation of D_m is

$$E(D_m) = J(1 - e^{-2vt}) \quad , \tag{4}$$

where J is the expected homozygosity and is assumed to be constant throughout the evolutionary process (Li and Nei 1975). The sampling and drift variances of D_m have also been studied (Nei and Roychoudhury 1974; Li and Nei 1975). It is clear from (4) that the relationship between D_m and t is nonlinear but when $t \ll 1/(2v)$ it is approximately linear, i.e., $E(D_m) \approx 2vJt$. Therefore, when closely related populations are to be studied, this measure can also be used for making a phylogenetic tree.

Another measure that is often used for molecular taxonomy is Rogers' distance. This is defined as

$$D_R = \frac{1}{r} \sum_{j=1}^{r} \left[\sum_{i=1}^{m_j} (x_{ij} - y_{ij})^2/2 \right]^{1/2} . \quad (5)$$

Some authors prefer this measure, because this satisfies the principle of triangle inequality. We note that Farris' (1972) method of tree making was originally developed by using a metric that satisfies this principle. It is therefore interesting to examine whether or not this measure gives a better tree when Farris' method is used. No analytical study has been made about the expected relationship between D_R and divergence time, but Nei's (1976) computer simulation has shown that as t increases D_R rapidly increases in the early stage of evolution but the rate of increase of D_R gradually declines as in the case of D_m. D_R has an upper bound, which is usually smaller than 1 when r is large.

Some human geneticists have used Cavalli-Sforza's (1969) measure f_θ after he showed that it has similaritiy with Wright's F_{ST} when there are only two alleles at a locus. It is defined as

$$f_\theta = \frac{4}{r} \sum_{j=1}^{r} \left(1 - \sum_{i=1}^{m_j} \sqrt{x_{ij}y_{ij}}\right) / (m_j - 1) . \quad (6)$$

Cavalli-Sforza (1969) conjectured that the expectation of f_θ is given by $1 - e^{-t/2N}$, where N is the effective population size of populations X and Y. Nei (1976) showed that the relationship between f_θ and t is not as simple as Cavalli-Sforza's speculation but in the early stage of evolution f_θ again increases almost linearly with time. Nei (unpublished) also noted that this measure has a smaller coefficient of variation than Nei's standard distance (D) when D or f_θ is small. It is therefore possible that when a phylogenetic tree is constructed for closely related populations f_θ gives a better tree than D does.

However, Cavalli-Sforza's f_θ has one deficiency; it is heavily dependent on the number of low-frequency alleles in the sample, though these alleles do not contribute very much to the average gene differences between populations. It is known that the number of low-frequency alleles increases substantially with increasing sample size, and thus f_θ is expected to decrease with increasing sample size even for the same pair of populations (see formula (6)). This deficiency can be removed if we use the following measure.

$$D_A = \frac{1}{r} \sum_{j=1}^{r} \left(1 - \sum_{i=1}^{m_j} \sqrt{x_{ij}y_{ij}}\right) . \quad (7)$$

This has a close relationship with Bhattacharrya's (1946) angular transformation, and it can be shown that D_A is proportional to Sanghvi's (1953) distance when the distance is small. It does not depend on the number of low-frequency alleles very much, since m_j in the denominator of (6) has been eliminated.

In the present paper we have used the above five distance measures to study the accuracies and efficiencies of the three tree-making methods mentioned earlier. One (D_R) of the distance measures is a metric, but the others are not unless a very large number of loci are used. As will be seen later, the accuracies and efficiencies of the tree-making methods depend on the distance measure used to a considerable extent.

2) Methods of Simulation. As in the previous paper, we considered 8 populations or OTUs (operational taxonomic units), and these OTUs were assumed to evolve following the model tree in Fig. 3a of Tateno et al. (1982). (This model tree is known to give more errors in the reconstructed tree than the model tree in Tateno et al.'s Fig. 3b, but the relative accuracies of the three tree-making methods are nearly the same for both model trees.) When the expected number of gene substitutions (M) for the shortest branch in the model tree is 0.1, it becomes Fig. 1a of this paper. Unlike our previous study, however, we followed the changes of allele frequencies in each population in the entire evolutionary process rather than the changes of nucleotide sequences by using computer simulation. In this simulation the allele frequency changes were assumed to occur by mutation and random genetic drift, ignoring the effect of selection. In each generation mutations were introduced at a specified rate, and the genes for the next generation were sampled at random. The mutations introduced were always new and different from the extant alleles. When a population split into two, we assumed that two identical populations were produced in a particular generation and immediately reproductive isolation was generated. In practice, there would generally be some migration at least for a while after two populations are separated, but the effect of this migration is known to be small unless the evolutionary time considered is extremely short (Li 1976). We also assumed that the population size (N) is the same for all populations.

In the present case the expected number of gene substitutions (M in Fig. 3a of Tateno et al. 1982) for the unit evolutionary time (the shortest distance), t_0, is given by vt_0. Note that Nei's standard genetic distance measures

the number of gene substitutions, so that the expected distance between OTUs 1 and 2, for example, is $D \equiv 2M = 2vt_0$ from (2). We used two different values of M, i.e., 0.004 and 0.1. The case of M = 0.004 roughly corresponds to the differentiation of populations within a species, whereas M = 0.1 corresponds to the differentiation of species within a genus (see Nei 1975, pp. 184–185). In both cases of M = 0.1 and M = 0.004 we used 4Nv = 0.2, which corresponds to an average heterozygosity of $H \equiv 4Nv/(1 + 4Nv) = 0.167$. We note that in many Drosophila species the average heterozygosity is of this order of magnitude.

In the case of M = 0.004 we simulated the evolutionary changes of gene frequencies using the Monte Carlo method described by Nei and Tateno (1975). From the study of Li and Nei (1975), it is known that the mean and variance of D for a given number of loci are determined by Nv rather than by N and v separately. We therefore used a high mutation rate (v = 0.002) and a small effective size (N = 25) to save computer time. Actually, we examined the mean and variance of D for the case of v = 0.002 and N = 50 in addition to the above case, but the results obtained were virtually the same as those for v = 0.002 and N = 25. The initial allele frequencies in the ancestral population were generated by using Griffiths and Li's (1983) computer algorithm. At the end of the evolutionary process, allele frequencies were recorded for all populations, and the five genetic distances mentioned above were computed for each pair of populations.

In the case of M = 0.1 the allele frequencies in the final populations were generated by using Griffiths and Li's (1983) computer algorithm. This algorithm is based on Griffiths' (1980) theoretical work on the transient distribution of allele frequencies and saves computer time to a great extent when t is large. Unfortunately, this algorithm does not give accurate results when vt is extremely small. Griffiths' theoretical distribution of allele frequencies is a function of Nv and t only. Therefore, the results obtained for given values of Nv and t are applicable to any population size or any mutation rate as long as Nv remains the same. In practice, however, we must know the number of alleles in the sample to compute genetic distances. To specify this number we assumed that the alleles of which the population frequency is equal to or higher than 0.001 are observable in the sample and m_j is equal to the number of alleles satisfying this condition in the two populations concerned. This is equivalent to using a sample of 500 diploid individuals from a large population.

In both cases of M = 0.1 and M = 0.004, gene frequency data were generated for 100 loci in each replication, and this was repeated 10 times. In each replication the five distance measures mentioned earlier were computed by using gene frequency data for the first 10 loci, first 20 loci, first 30 loci,..., and all 100 loci to see the effect of the number of loci on reconstructed trees. Therefore, 50 distance matrices for 8 OTUs were obtained in each replication. Both rooted and unrooted trees were reconstructed by the three tree-making methods for each of these distance matrices. Thus, the total number of trees reconstructed in this study was 6000. The most ancestral point (root) for the trees reconstructed by the Farris and modified Farris methods was determined by assuming that the evolutionary rates of the two most divergent OTUs are equal (cf. Tateno et al. 1982).

As discussed by Teteno et al. (1982), there are two criteria for measuring the deviation of a reconstructed tree from the model tree. One is the degree of distortion of the topology of the reconstructed tree, and the other is the amount of deviation of patristic (estimated) branch lengths from true lengths. To measure the topological errors, we used Robinson and Foulds' (1981) distortion index (d_T), which is roughly twice the number of interchanges of OTUs required for converting the topology of a reconstructed tree to that of the true tree (see Tateno et al. 1982). When the topology of a reconstructed tree is correct, d_T takes a value of zero. In addition to d_T we also examined the proportion of replications in which the correct topology was obtained. To measure the errors in the estimates of branch lengths, we used Tateno et al.'s (1982) measures, i.e., the average deviation of patristic distances from the expected distances (S_E) and the average deviation of patristic distances from the observed distances (S_O). Here the patristic distance between OTUs i and j refers to the distance that is obtained by summation of the lengths of all branches linking this particular pair of OTUs in the reconstructed tree, whereas the expected distance is the corresponding distance given by the model tree. The observed distance is simply the observed value between a pair of OTUs in the distance matrix.

Mathematically, S_E and S_O are defined as

$$S_E = \left[2 \sum_{i>j} (D_{ij} - D'_{ij})^2/(n(n-1))\right]^{1/2}, \quad (8)$$

$$S_O = \left[2 \sum_{i>j} (D_{ij} - D''_{ij})^2/(n(n-1))\right]^{1/2}, \quad (9)$$

where D_{ij}, D'_{ij}, and D''_{ij} are the patristic distance, expected distance, and observed distance between OTUs i and j, respectively, and n is the number of OTUs.

When a small number of loci are used, the observed number of gene substitutions per locus for a given branch can be quite different from the expected number because of stochastic errors, even if the expected rate of gene substitution is constant. The tree that is subject to this type of errors or to truly varying rates of substitution is called gene-tree (Nei 1977, Tateno et al. 1982). Gene-tree will vary with the gene or group of genes used, even though the true genealogical tree must be the same for the entire set of genes in the genome. On the other hand, the tree that describes the actual evolutionary pathways of the OTUs in question is called species-tree.

Most evolutionists are interested in knowing species-tree rather than gene-tree. Of course, as the number of loci used increases, gene-tree is expected gradually to converge to species-tree (or genome-tree), if the average rate of gene substitution per year is constant. In practice, the number of loci used is rather limited at the present time, but it is still possible to estimate a species-tree. Nei (1975) noted that species-tree can be estimated by UPGMA even if the number of loci used is not very large. On the other hand, the Farris and modified Farris methods are intended to estimate a gene-tree.

As is clear from their definitions, S_E is intended to measure the deviation from the species-tree, whereas S_O the deviation from the gene-tree.

Results

Topological Errors

1) Case of M = 0.1. Table 1 shows the expected numbers of gene substitutions per locus (expected distances) and the observed values of D's for all pairs of the eight OTUs for the case of M = 0.1 in one replication (replication 3) of computer simulation. These distances are based on gene frequency data for 50 loci. It is clear that despite the large number of loci used the observed value of D is appreciably different from the expected value for some pairs of OTUs such as 1 and 6 and 3 and 4. This is of course due to the stochastic nature of mutation and gene frequency change in the evolutionary process. When the number of loci used is small, the deviations of the observed distances from the expected values are even larger, and the deviations gradually decline as the number of loci increases. However, our computer simulation has shown that even with 100 loci the deviations are not negligible.

Because of this stochastic error the evolutionary tree reconstructed is usually considerably different from the model tree given in Figure 1a. Figures 1b, 1c, and 1d show the trees reconstructed by UPGMA, the Farris method, and the modified Farris method, respectively, by using the observed values of D. The trees reconstructed by UPGMA have the same topology as that of the true tree, though there are some errors in the estimates of branch lengths. Therefore, the distortion index for this tree is 0. The topologies of the trees reconstructed by the other two methods are both incorrect. The d_T values for the Farris and modified Farris methods are 8 and 2, respectively. Table 2 gives the patristic distances obtained by UPGMA and the modified Farris method. The patristic distances obtained by UPGMA tend to be close to the expected distances in Table 1, whereas those obtained by the modified Farris method tend to be close to the observed distances. Nevertheless, the differences between the patristic and expected distances or between the patristic and observed distances are substantial in some pairs of OTUs. S_E and S_O are respectively 0.073 and 0.109 for UPGMA and 0.138 and 0.106 for the modified Farris method.

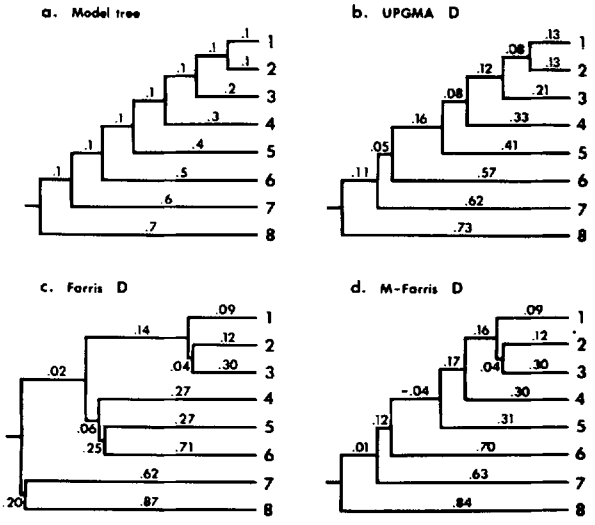

Fig. 1. Model tree and reconstructed trees (b, c, d) by using D in replication 3 of computer simulation. The value given to each branch (or internode) in model tree is the expected number of gene substitutions, whereas the corresponding number in reconstructed tree is the estimate of branch length. 4Nv = 0.2, M = 0.1, and the number of loci used is 50. (b) $d_T = 0$, $S_E = 0.073$, $S_O = 0.109$. (c) $d_T = 8$, $S_E = 0.250$, $S_O = 0.145$. (d) $d_T = 2$, $S_E = 0.138$, $S_O = 0.106$

Table 1. Observed values of genetic distances based on gene frequency data for 50 loci in one replication (replication 3) of computer simulation (above the diagonal) and expected numbers of gene substitutions (below the diagonal). Nei's distance (D) is used. 4Nv = 0.2, and M = 0.1

OTU	1	2	3	4	5	6	7	8
1		0.260	0.434	0.556	0.805	1.249	1.063	1.297
2	0.2		0.423	0.618	0.803	1.139	1.083	1.259
3	0.4	0.4		0.808	0.904	1.200	1.291	1.556
4	0.6	0.6	0.6		0.786	1.123	1.167	1.422
5	0.8	0.8	0.8	0.8		0.978	1.340	1.567
6	1.0	1.0	1.0	1.0	1.0		1.459	1.650
7	1.2	1.2	1.2	1.2	1.2	1.2		1.487
8	1.4	1.4	1.4	1.4	1.4	1.4	1.4	

Table 2. Patristic distances obtained by UPGMA (below the diagonal) and the modified Farris method (above the diagonal) from observed distances in Table 1. D is used. 4Nv = 0.2 and M = 0.1

OTU	1	2	3	4	5	6	7	8
1		0.260	0.434	0.556	0.732	1.086	1.137	1.362
2	0.260		0.423	0.626	0.803	1.156	1.207	1.432
3	0.429	0.429		0.800	0.977	1.330	1.381	1.606
4	0.661	0.661	0.661		0.786	1.139	1.190	1.415
5	0.824	0.824	0.824	0.824		0.978	1.029	1.254
6	1.138	1.138	1.138	1.138	1.138		1.459	1.684
7	1.234	1.234	1.234	1.234	1.234	1.234		1.487
8	1.463	1.463	1.463	1.463	1.463	1.463	1.463	

Table 3. Observed distances (below the diagonal) and patristic distances (above the diagonal) obtained by UPGMA when D_R is used. The same gene frequency data as those for Tables 1 and 2 are used. 4Nv = 0.2 and M = 0.1

OTU	1	2	3	4	5	6	7	8
1		0.313	0.399	0.499	0.565	0.654	0.679	0.732
2	0.313		0.399	0.499	0.565	0.654	0.679	0.732
3	0.398	0.400		0.499	0.565	0.654	0.679	0.732
4	0.453	0.487	0.559		0.565	0.654	0.679	0.732
5	0.556	0.559	0.597	0.550		0.654	0.679	0.732
6	0.678	0.648	0.676	0.650	0.620		0.679	0.732
7	0.637	0.636	0.693	0.665	0.713	0.733		0.732
8	0.692	0.682	0.757	0.722	0.761	0.768	0.742	

Table 4. Observed distances (below the diagonal) and patristic distances (above the diagonal) obtained by UPGMA when f_θ is used. The same gene frequency data as those for Tables 1 and 2 are used. 4Nv = 0.2 and M = 0.1

OTU	1	2	3	4	5	6	7	8
1		0.381	0.535	0.650	0.755	0.846	0.885	0.978
2	0.381		0.535	0.650	0.755	0.846	0.885	0.978
3	0.556	0.515		0.650	0.755	0.846	0.885	0.978
4	0.596	0.615	0.740		0.755	0.846	0.885	0.978
5	0.754	0.724	0.815	0.725		0.846	0.885	0.978
6	0.869	0.822	0.898	0.822	0.818		0.885	0.978
7	0.831	0.817	0.945	0.845	0.933	0.938		0.978
8	0.937	0.903	1.036	0.940	1.021	1.012	0.999	

Table 3 gives the D_R values obtained from the same gene frequency data as those used for computing the D values in Table 1 and the patristic distances obtained by UPGMA. It is clear that D_R is greater than the expected number of gene substitutions when this number is small but smaller than the expected number when this is large. Thus, when the expected number or expected distance is 0.2 (between OTUs 1 and 2), D_R is 0.313, and when the expected distance is 1.4 (between OTU 8 and the other OTUs), it is 0.68 to 0.77. Clearly, D_R is not proportional to the number of gene substitutions as indicated by Nei (1976) (see also Figure 5). Furthermore, the patristic distances for the tree obtained by UPGMA are not necessarily close to the observed distances. A similar pattern is also observed for Cavalli-Sforza's distance f_θ, though this quantity can be greater than 1 (Table 4).

Figure 2 shows the evolutionary trees constructed by the three tree-making methods with D_R and f_θ for the same set of gene frequency data (replication 3). UPGMA again gives the correct topology for both D_R and f_θ, but the other tree-making methods produce an incorrect topology. To compare the efficiencies of different tree-making methods, however, we must examine the average performances of the methods over all replications.

Rooted Trees. Table 5 shows the proportion (P) of replications in which the correct topology was obtained. Each value in this table is based on ten reconstructed

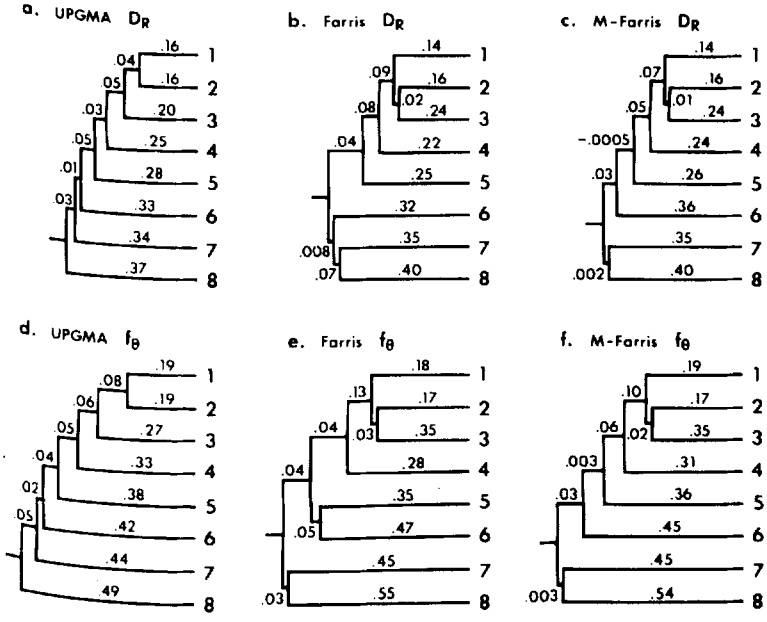

Fig. 2. Reconstructed trees by using D_R and f_θ in replication 3 of computer simulation. The value given to each branch is the estimate of branch length. 4Nv = 0.2, M = 0.1, and the number of loci used is 50. (a) $d_T = 0$, $S_E = 0.448$, $S_0 = 0.030$. (b) $d_T = 6$, $S_E = 0.388$, $S_0 = 0.086$. (c) $d_T = 4$, $S_E = 0.449$, $S_0 = 0.032$. (d) $d_T = 0$, $S_E = 0.270$, $S_0 = 0.045$. (e) $d_T = 6$, $S_E = 0.243$, $S_0 = 0.068$. (f) $d_T = 4$, $S_E = 0.278$, $S_0 = 0.038$

Table 5. The proportion of replications in which the correct topology was obtained. 4Nv = 0.2 and M = 0.1

No. of loci	UPGMA					Farris					Modified Farris				
	D	D_m	D_R	f_θ	D_A	D	D_m	D_R	f_θ	D_A	D	D_m	D_R	f_θ	D_A
						Rooted tree									
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0.1	0	0	0.1	0	0	0	0	0	0.1	0	0.2	0.1	0.1
30	0.1	0.1	0.1	0.1	0.2	0	0	0	0	0	0.2	0.2	0	0.2	0.1
40	0.1	0.1	0.1	0.3	0.1	0.1	0.1	0	0.1	0.1	0.2	0.3	0.2	0.2	0.3
50	0.2	0.2	0.2	0.3	0.2	0.2	0.1	0.1	0.2	0	0.3	0.3	0.3	0.2	0.3
60	0.2	0.3	0.2	0.3	0.3	0.1	0.1	0	0.1	0.1	0.3	0.4	0.3	0.4	0.3
70	0.4	0.4	0.4	0.4	0.5	0.2	0	0	0.1	0	0.4	0.4	0.3	0.4	0.3
80	0.4	0.5	0.4	0.3	0.6	0.2	0.1	0.1	0.1	0.2	0.6	0.7	0.5	0.5	0.9
90	0.5	0.5	0.4	0.4	0.6	0.3	0.1	0.1	0.2	0.2	0.7	0.7	0.7	0.6	0.9
100	0.6	0.7	0.6	0.3	0.7	0.5	0.3	0.3	0.3	0.5	0.8	0.8	0.8	0.6	0.9
						Unrooted tree									
10	0	0	0	0	0	0	0.2	0.2	0.1	0.2	0	0.2	0.2	0.2	0.2
20	0.1	0.1	0	0	0.2	0.2	0.2	0.3	0.4	0.3	0.3	0.3	0.4	0.5	0.5
30	0.1	0.1	0.1	0.2	0.2	0.2	0.2	0.3	0.4	0.3	0.3	0.6	0.4	0.4	0.4
40	0.1	0.1	0.1	0.3	0.1	0.2	0.3	0.2	0.3	0.4	0.3	0.4	0.3	0.3	0.5
50	0.2	0.2	0.2	0.3	0.2	0.3	0.5	0.4	0.5	0.4	0.5	0.4	0.4	0.5	0.5
60	0.2	0.3	0.2	0.3	0.3	0.3	0.5	0.4	0.5	0.5	0.5	0.4	0.5	0.6	0.5
70	0.4	0.4	0.4	0.4	0.5	0.3	0.4	0.5	0.5	0.5	0.6	0.6	0.6	0.6	0.6
80	0.4	0.5	0.4	0.5	0.6	0.5	0.5	0.5	0.8	0.7	0.7	0.9	0.8	1.0	1.0
90	0.5	0.5	0.4	0.4	0.6	0.6	0.7	0.6	0.8	0.7	0.8	0.9	0.9	1.0	1.0
100	0.6	0.7	0.6	0.3	0.7	0.9	0.9	0.9	1.0	1.0	0.9	1.0	1.0	1.0	1.0

trees. When rooted trees are constructed by using 10 loci, P is 0 for all tree-making methods irrespective of the distance measures used. It increases as the number of loci (r) increases, but even with 100 loci it is not 100 percent. In general, UPGMA and the modified Farris method show a better performance than the Farris method with this criterion. The poor performance of the Farris method seems to be due to overestimation of branch lengths that occasionally occur in this method (Tateno et al. 1982). This overestimation is caused by chance effects, and it is larger in long branches than in short branches. Furthermore, a large part of topological errors in this method occurs when the root is given to a reconstructed tree, as will be seen later.

The differences in P among different distance measures are rather small, but D and D_A tend to give a higher proportion of correct topologies than the other distance measures. As mentioned earlier, Farris (1981)

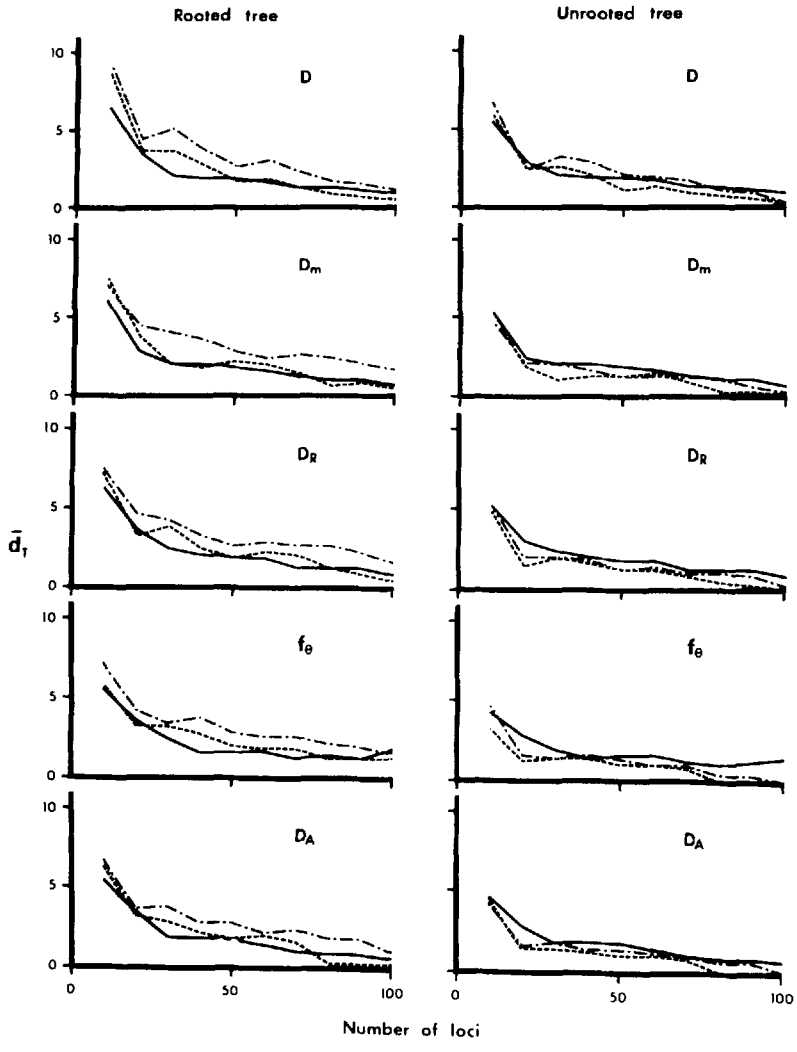

Fig. 3. Relationships between $\bar{d}_T$ and the number of loci used for the case of M = 0.1. *Solid line:* UPGMA. *Chain line:* Farris method. *Broken line:* Modified Farris method

criticized the use of D for the reason that it is not a metric, and Swofford (1981) suggested that D_R be used in conjunction with the Farris method. Table 5 indicates that for obtaining the correct topology D_R shows a poor performance compared with the other distance measures in any tree-making method. Indeed, it is the poorest distance measure when the Farris method is used. Clearly Farris' or Swofford's contention is not supported.

The average distortion indices ($\bar{d}_T$) over ten replications for the case of M = 0.1 are given in Fig. 3 in relation to the number of loci used. It is seen that the $\bar{d}_T$ for rooted trees is very large when r = 10 but rapidly declines as r increases. However, the decrease of $\bar{d}_T$ with increasing r is nonlinear, and the rate of decrease in $\bar{d}_T$ is not very large when r is equal to or larger than 30. When 30 loci are used, $\bar{d}_T$ is already about 2 in UPGMA, which means that the amount of error of the reconstructed tree is about one interchange of branches from the true tree. As the number of loci increases further, $\bar{d}_T$ decreases very slowly, and even with r = 100, $\bar{d}_T$ is not 0. It is also seen that UPGMA and the modified Farris method generally give a smaller value of $\bar{d}_T$ than the Farris method for all values of r. This is so even when D_R is used as a distance measure. This again refutes Swofford's contention. The difference in $\bar{d}_T$ between UPGMA and the modified Farris method is small, but the latter tends to show a little better performance than the former when the number of loci is large. The differences among different distance measures are also small, though D_A tends to give a better topology than the other measures. There is a strong negative correlation between P and $\bar{d}_T$, as expected.

Unrooted Trees. Most evolutionists are interested in making a rooted tree, but under certain circumstances an unrooted tree is constructed. The topological errors for unrooted trees are equal to or smaller than those for rooted trees, because in rooted trees an additional error may be generated in the process of putting the root. The difference in topological errors between rooted and unrooted trees depends on the tree-making method; it is

much larger in the Farris and modified Farris methods than in UPGMA. Thus, the proportion (P) of correct topologies obtained for unrooted trees is considerably higher in the former two methods than in UPGMA for any number of loci (Table 5). In this case even 10 loci give the correct topology with a frequency of 20 percent. The $\bar{d}_T$ value is also smaller in unrooted trees than in rooted trees and when the number of loci is 80 – 100, $\bar{d}_T$ is often 0 or close to 0 (Fig. 3). In terms of $\bar{d}_T$ the modified Farris method generally shows the best performance in topology-making irrespective of the distance measure used. These results are consistent with those obtained with nucleotide substitution in long-term evolution (Tateno et al. 1982). The Farris method is generally better than UPGMA when D_m, D_R, f_θ, and D_A are used, but when D is used, it is worse than the latter except when the number of loci used is 80 or larger.

2) Case of M = 0.004. When M is very small, genetic distance has a large coefficient of variation (Li and Nei 1975), so that the topology of a reconstructed tree is expected to be subject to a large stochastic error. Table 6 shows the proportion (P) of correct topologies among the ten replications for the case of M = 0.004. Comparison of this table with Table 5 indicates that P is substantially lower for M = 0.004 than for M = 0.1 in both rooted and unrooted trees. Particularly when a rooted tree is constructed with r equal to 30 or less, the correct topology is rarely obtained. In the case of rooted trees, the Farris method again shows the poorest performance, and when D_R is used no correct topology has been obtained for any number of loci used. The modified Farris method showed a slightly better performance than the Farris method particularly when f_θ and D_A are used. UPGMA again shows the best result when P is used as a criterion. Among the five distance measures examined, f_θ and D_A, particularly the latter measure, show a good performance in all the tree-making methods.

The values of $\bar{d}_T$ for M = 0.004 are presented in Figure 4. This quantity is always smaller in the trees reconstructed by UPGMA than in those reconstructed by the other two methods, and the $\bar{d}_T$ value for the modified Farris method is generally smaller than that for the Farris method. Therefore, the results obtained from this criterion are virtually the same as those obtained from P.

As mentioned earlier, Tateno (1982) studied the topological errors of reconstructed trees by using his own distortion index DI but did not find much difference between the tree-making methods. He used only Nei's distance D for the case of M = 0.002. A close examination of his Fig. 4, however, indicates that DI tends to be smaller in UPGMA than in the Farris method. The small difference between the two methods seems to be due to the inefficiency of his index. Although there is a high correlation between d_T and DI when d_T is relatively small (Tateno et al. 1982), DI is not proportional to the number of branch interchanges between the two trees compared. This non-proportionality seems to have reduced the power to

Table 6. The proportion of replications in which the correct topology was obtained. 4Nv = 0.2 and M = 0.004

No. of loci	UPGMA					Farris					Modified Farris				
	D	D_m	D_R	f_θ	D_A	D	D_m	D_R	f_θ	D_A	D	D_m	D_R	f_θ	D_A
						Rooted tree									
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1
40	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0.1	0	0	0	0.1	0
50	0.1	0.1	0.2	0.1	0.2	0	0	0	0	0.1	0	0	0.1	0	0.1
60	0.1	0.1	0	0.3	0.3	0	0	0	0	0.1	0.1	0	0.2	0.3	0.5
70	0.1	0.1	0.1	0.4	0.3	0	0	0	0	0.1	0	0	0	0.1	0.2
80	0.1	0.1	0.1	0.4	0.3	0	0	0	0.2	0.1	0	0	0	0.2	0.3
90	0	0	0.2	0.5	0.4	0	0	0	0.2	0.1	0	0	0	0.1	0.2
100	0.2	0.2	0.2	0.3	0.6	0.1	0.1	0	0.1	0	0	0	0	0.2	0.2
						Unrooted tree									
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0
30	0	0	0	0	0	0	0	0	0.2	0.1	0.1	0.1	0	0.1	0.1
40	0.1	0.1	0.2	0.2	0.1	0	0	0.1	0.2	0.3	0.1	0.1	0.2	0.2	0.2
50	0.1	0.1	0.2	0.1	0.2	0.2	0.1	0.1	0.2	0.2	0.2	0.1	0.2	0.1	0.3
60	0.1	0.1	0	0.3	0.3	0.2	0.1	0	0.2	0.2	0.2	0.2	0.3	0.3	0.5
70	0.3	0.3	0.1	0.5	0.3	0.1	0.1	0.1	0.5	0.5	0	0	0	0.2	0.4
80	0.2	0.2	0.3	0.5	0.3	0.1	0.1	0.1	0.6	0.5	0.1	0.1	0.1	0.5	0.3
90	0.1	0.1	0.3	0.6	0.4	0.1	0.1	0.1	0.6	0.4	0.1	0.1	0	0.5	0.3
100	0.3	0.3	0.4	0.4	0.7	0.3	0.4	0	0.4	0.3	0.2	0.2	0	0.5	0.3

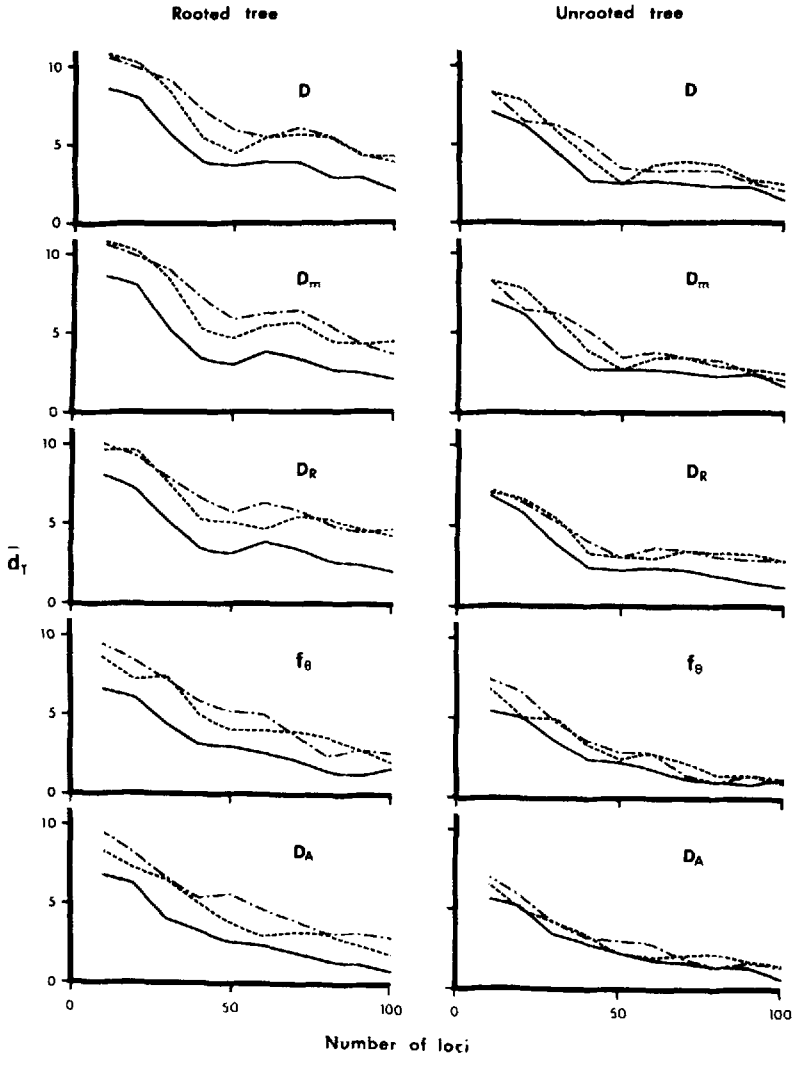

Fig. 4. Relationships between $\bar{d}_T$ and the number of loci used for the case of M = 0.004. *Solid line:* UPGMA. *Chain line:* Farris method. *Broken line:* Modified Farris method

detect the difference in accuracy of topology construction between different methods. (Furthermore, there were some errors in his construction of topologies.)

Figure 4 shows that, like the case of M = 0.1, D_A and f_θ give a smaller value of d_T than the other distance measures. The better performance of f_θ compared with D when genetic distance is small can be explained by the fact that the coefficient of variation of f_θ is smaller than that of D. To see whether or not all the results in Table 6 can be explained by this factor alone, we computed the coefficients of variation for all distance measures (observed values). Table 7 shows the results for the case of 60 loci; the results for other numbers of loci are not presented because they showed essentially the same pattern as that for 60 loci. This table indicates that f_θ and D_A indeed have a smaller coefficient of variation than D and D_m, but D_R, which has shown a rather poor performance in topology making, also has a small coefficient. Therefore, the coefficient of variation alone is not sufficient for explaining our results. Actually the quantity which is more directly related to P or $\bar{d}_T$ is the coefficient of variation of the estimates of the shortest branches (internodes) of the tree in Figure 1a. Indeed, a distance measure that has a relatively large value of P or a small value of $\bar{d}_T$ in Table 6 tends to show small coefficients of variation of the estimates of the shortest branches ($D_{ij} - D_{kl}$ in Table 7). Therefore, a small value of P or a large value of $\bar{d}_T$ is apparently caused by random errors of the differences between distance estimates that correspond to the shortest branches.

The results for unrooted trees for the case of M = 0.004 are somewhat different from those for the case of M = 0.1. Namely, unlike the case of M = 0.1, UPGMA generally shows a better performance in topology making than the other two methods even for unrooted trees. This is clear from the values of P (Table 6) and $\bar{d}_T$ (Fig. 4). On the other hand, the Farris and modified Farris methods show more or less the same values of P and $\bar{d}_T$ in all distance measures. As in the case of rooted trees, D_A and f_θ generally show a larger value of P and a smal-

Table 7. Coefficients of variation of various distance measures. D_{ij} refers to the distance between OTUs i and j, whereas $D_{ij}-D_{k\ell}$ is the difference between distances D_{ij} and $D_{k\ell}$. 4Nv = 0.2, M = 0.004, and the number of loci used is 60

	D	D_m	D_R	f_θ	D_A
D_{12}	0.118	0.097	0.084	0.078	0.121
D_{34}	0.302	0.296	0.113	0.091	0.122
D_{56}	0.249	0.226	0.137	0.111	0.176
D_{78}	0.292	0.280	0.172	0.126	0.170
$D_{34}-D_{12}$	0.472	0.467	0.291	0.187	0.216
$D_{45}-D_{23}$	0.406	0.383	0.411	0.253	0.327
$D_{56}-D_{34}$	0.657	0.608	0.682	0.373	0.535
$D_{67}-D_{45}$	1.229	1.115	1.775	0.663	0.776
$D_{78}-D_{56}$	1.499	1.478	1.538	0.724	0.895

ler value of $\bar{d}_T$ than the other distance measures in all tree-making methods.

The effect of the number of loci used on the accuracy of topology for the case of M = 0.004 is also slightly different from that for the case of M = 0.1. In the case of M = 0.1 the accuracy increased rapidly as the number of loci increased from 10 to 30, but after r = 30 the increase of the accuracy was slow. In the case of M = 0.004 the accuracy for rooted trees increases almost linearly until r reaches about 40, and then the rate of increase is generally slowed down. When trees are made by UPGMA with D_A, however, $\bar{d}_T$ decreases almost linearly even up to r = 100. Essentially the same pattern is observed for unrooted trees.

Errors of the Estimates of Branch Lengths

1) Case of M = 0.1. Another important criterion of the accuracy of a reconstructed tree is the deviation of estimates of branch lengths from true branch lengths. We have seen that the topology of the tree reconstructed by UPGMA in replication 3 for M = 0.1 is correct irrespective of the distance measure used (Figs. 1b, 2a, and 2d). However, the estimates of branch lengths are considerably different from each other. Comparison of these trees with the true tree. (Fig. 1a) indicates that D gives a better result for estimating branch lengths than other distance measures. Indeed, the S_E value for D is 0.073, whereas the S_E's for D_R and f_θ are 0.448 and 0.270, respectively. Therefeore, with this criterion the tree produced by UPGMA with use of D is the best among the three. One might think that the Farris method gives a good tree when D_R is used. That this is not the case can be seen from comparison of Fig. 2b and Fig. 1b. Compared with D, D_R generally gives a tree in which the part near the root is condensed whereas the other part is elongated. This is because D_R is not proportional to the expected number of gene substitutions. A similar pattern is observed for D_m, f_θ, and D_A, though the results for D_m and D_A are not shown here. In other replications D did not necessarily give a correct topology, but the properties of estimated branch lengths from the five distance measures were essentially the same.

The average values ($\bar{S}_E$) of S_E's over all replications for the cases of 20 loci, 60 loci, and 100 loci examined are presented in Table 8. Since the general trends of the effect of number of loci can be seen from these cases, $\bar{S}_E$'s for other numbers of loci are not included. The value of $\bar{S}_E$ varies considerably with the tree-making method and the distance measure used. The smallest value is obtained when UPGMA with D is used. This supports our visual conclusion from Fig.1 and 2. When D is used, the modified Farris method also shows a relatively small value of $\bar{S}_E$. In contrast, the $\bar{S}_E$ for the Farris method is nearly twice as large as that of UPGMA. This large value of $\bar{S}_E$ is apparently caused by overestimation of branch lengths (Tateno et al. 1982, Table 9).

Table 8. Means of average deviations of patristic distances from expected distances ($\bar{S}_E$). 4Nv = 0.2 and M = 0.1. These results are based on 10 replications. All values should be divided by 10^3

	D	D_m	D_R	f_θ	D_A
			20 loci		
UPGMA	252 ± 26	539 ± 12	457 ± 11	305 ± 17	426 ± 13
Farris	461 ± 66	497 ± 13	417 ± 13	272 ± 16	386 ± 14
Modified Farris	291 ± 31	539 ± 11	458 ± 10	309 ± 16	427 ± 12
			60 loci		
UPGMA	136 ± 10	540 ± 5	456 ± 4	295 ± 11	426 ± 5
Farris	225 ± 12	511 ± 7	430 ± 5	268 ± 12	397 ± 5
Modified Farris	161 ± 9	541 ± 5	457 ± 4	297 ± 10	427 ± 5
			100 loci		
UPGMA	122 ± 8	534 ± 5	452 ± 4	296 ± 8	420 ± 5
Farris	204 ± 13	510 ± 5	427 ± 5	271 ± 9	395 ± 6
Modified Farris	140 ± 6	535 ± 5	452 ± 4	297 ± 8	421 ± 5

In the case of D, $\bar{S}_E$ decreases as the number of loci used increases, as expected.

When the other distance measures are used, UPGMA no longer gives the smallest value of $\bar{S}_E$. On the contrary, this method and the modified Farris method give a larger value of $\bar{S}_E$ than the Farris method. The smallest value of $\bar{S}_E$ for the Farris method when D_m, D_R, f_θ, and D_A are used is apparently due to the nonlinear relationship of these distance measures with the number of gene substitutions and overestimation of branch lengths in this method. As mentioned earlier, the values of these distance measures are usually smaller than the number of gene substitutions when the latter is larger than about 0.6 (Tables 1, 3, and 4), whereas the Farris method tends to give overestimates for long branches. Therefore, when the Farris method is used, the differences between patristic distances and expected distances are reduced to some extent. When D_m, D_R, f_θ, and D_A are used, $\bar{S}_E$ does not necessarily decrease with increasing r; rather it remains nearly the same irrespective of the number of loci used. This is due to the fact that the expectations of these distance measures are not equal to the branch lengths of the true tree, i.e., the expected numbers of gene substitutions, and thus even if the effect of stochastic errors are reduced, $\bar{S}_E$ does not necessarily decrease.

The fact that the Farris method indeed gives overestimates of long branches in the present case can be seen from Table 9, where the observed values of D and the patrisitic distances for some pairs of OTUs are presented. It is clear that the patristic distances obtained by UPGMA and the modified Farris method are close to the observed distances and none of the differences is statistically significant. In the case of the Farris method, however, the patristic distances are generally greater than the observed values particularly when the distance is large. This result corrorates Tateno et al.'s (1982) similar finding with the evolutionary trees reconstructed from nucleotide sequence data. A similar tendency was also observed for the case of M = 0.004 in the present study. At this point it should be mentioned that overestimation of branch lengths in the Farris method occurs by chance effects, so that the variance of the estimates is also ecpected to be large. Comparison of the standard errors of the estimates in Table 9 among the tree-making methods indicates that this is indeed the case.

In Table 8 f_θ gives a smaller value of $\bar{S}_E$ than the other distance measures except D. This is caused by larger patristic distances for f_θ for long branches except for D, as seen from Tables 1–4.

Although S_E is a quite good measure of the deviation of patristic branch lengths from true branches (Tateno et al. 1982), it cannot be computed for a tree constructed from actual data because in this case we do not know the true tree. For this reason the average deviation of patristic distances from observed distances (S_O) or a similar quantity has been used by many authors (e.g. Fitch and Margoliash 1967; Farris 1972; Prager and Wilson 1978). Tateno et al. (1982) have shown that this quantity is a rather poor measure of deviation of estimated branch lengths from true or actual branch lengths, but because there is no other measure computable they have suggested that S_O be used as a very crude measure. They have also noted that there is little correlation between S_O and S_E or between S_O and d_T.

At any rate, when D is used, the S_O values for the trees reconstructed by UPGMA, the Farris method, and

Table 9. Observed and patristic distances for some pairs of OTUs for the case of M = 0.1. The distance measure used is D. The results obtained are based on ten replications

OTUs	Observed distance	Patristic distance		
		UPGMA	Farris	M-Farris
		20 loci		
1 and 4	0.54 ± 0.04	0.54 ± 0.05	0.56 ± 0.05	0.54 ± 0.05
1 and 6	1.09 ± 0.10	1.12 ± 0.07	1.29 ± 0.17	1.08 ± 0.10
1 and 8	1.69 ± 0.17	1.56 ± 0.11	2.02 ± 0.17***	1.65 ± 0.11
		60 loci		
1 and 4	0.53 ± 0.04	0.54 ± 0.04	0.55 ± 0.04*	0.54 ± 0.04
1 and 6	1.00 ± 0.05	1.01 ± 0.04	1.10 ± 0.07*	1.00 ± 0.05
1 and 8	1.51 ± 0.08	1.54 ± 0.04	1.70 ± 0.08**	1.55 ± 0.06
		100 loci		
1 and 4	0.55 ± 0.03	0.56 ± 0.02	0.57 ± 0.03	0.55 ± 0.03
1 and 6	1.05 ± 0.05	1.05 ± 0.04	1.11 ± 0.06*	1.02 ± 0.05
1 and 8	1.56 ± 0.05	1.55 ± 0.03	1.77 ± 0.04**	1.56 ± 0.03

*, **, and *** indicate that the difference between the observed and patristic difference is significant at the 5%, 1%, and 0.1% level, respectively. Statistical tests were done by using the t-test based on pairwise comparisons rather than on the standard errors given in the table

the modified Farris method in replication 3 (in Figs. 1 and 2) are 0.109, 0.145, and 0.106, respectively. Therefore according to this criterion, the modified Farris method is the best and the Farris method is the poorest. Table 10 shows the $\bar{S}_O$ values over all replications for r = 20, 60, and 100 in the case of 4Nv = 0.2 and M = 0.1. Unlike $\bar{S}_E$, $\bar{S}_O$ declines with increasing number of loci for all distance measures. This is understandable because patristic distances are now compared with observed distances rather than with expected numbers of gene substitutions. When D is used, UPGMA no longer shows the best performance according to the criterion of S_Q; the best performance is attained by the modified Farris method. In other words, for estimating the branch lengths of a gene-tree the modified Farris method seems to be better than UPGMA. This result is the same as that for the case of nucleotide substitution (Tateno et al. 1982). The $\bar{S}_O$ values for other distance measures are smaller than for D because of their smaller absolute values for long branches. In all distance measures the modified Farris method gives the smallest value of $\bar{S}_O$, and the Farris method gives the largest value. The large value for the Farris method is caused by overestimation of branch lengths.

2) Case of M = 0.004. Table 11 shows the $\bar{S}_E$ for the case of M = 0.004. When D is used, the relative values of S_E for the three tree-making methods and the effect of number of loci are more or less the same as those for the case of M = 0.1, though the absolute values of S_E are much smaller in this case. In the other distance measures, however, the relative values of $\bar{S}_E$ are not the same as those for the case of M = 0.1. Unlike the previous case, the Farris method gives a larger value of $\bar{S}_E$ than other two methods in all cases but one. This is, however, again caused by overestimation of long branches in this

Table 10. Means of average deviations of patristic distances from observed distances ($\bar{S}_O$). 4Nv = 0.2 and M = 0.1. These results are based on 10 replications. All values should be divided by 10^3

	D	D_m	D_R	f_θ	D_A
			20 loci		
UPGMA	205 ± 26	43 ± 5	43 ± 4	59 ± 4	47 ± 5
Farris	268 ± 45	72 ± 8	73 ± 10	81 ± 9	73 ± 8
Modified Farris	136 ± 19	38 ± 4	39 ± 5	45 ± 5	37 ± 4
			60 loci		
UPGMA	97 ± 9	25 ± 2	24 ± 2	40 ± 1	27 ± 2
Farris	118 ± 10	44 ± 5	41 ± 4	49 ± 6	45 ± 6
Modified Farris	72 ± 6	22 ± 2	21 ± 1	26 ± 2	24 ± 2
			100 loci		
UPGMA	78 ± 6	19 ± 1	20 ± 1	30 ± 2	22 ± 1
Farris	118 ± 13	36 ± 4	36 ± 3	41 ± 4	37 ± 4
Modified Farris	60 ± 4	18 ± 1	18 ± 1	22 ± 2	19 ± 1

Table 11. Means of average deviations of patristic distances from expected distances ($\bar{S}_E$). 4Nv = 0.2 and M = 0.004. These results are based on 10 replications. All values should be divided by 10^4

	D	D_m	D_R	f_θ	D_A
			20 loci		
UPGMA	152 ± 13	136 ± 14	768 ± 41	1348 ± 47	237 ± 19
Farris	242 ± 18	179 ± 10	917 ± 41	1587 ± 64	351 ± 32
Modified Farris	170 ± 10	148 ± 11	783 ± 41	1365 ± 48	248 ± 20
			60 loci		
UPGMA	92 ± 10	90 ± 14	774 ± 35	1294 ± 19	188 ± 15
Farris	157 ± 15	108 ± 8	876 ± 40	1455 ± 24	253 ± 20
Modified Farris	103 ± 9	98 ± 12	778 ± 35	1303 ± 18	193 ± 15
			100 loci		
UPGMA	65 ± 9	84 ± 9	762 ± 29	1292 ± 18	179 ± 14
Farris	103 ± 15	77 ± 5	836 ± 32	1406 ± 26	224 ± 18
Modified Farris	77 ± 8	90 ± 8	765 ± 29	1297 ± 18	183 ± 14

Table 12. Means of average deviations of patristic distances from observed distances ($\bar{S}_0$). $4Nv = 0.2$ and $M = 0.004$ were used. These results are based on 10 replications. All values should be divided by 10^4

	D	D_m	D_R	f_θ	D_A
			20 loci		
UPGMA	103 ± 5	83 ± 4	134 ± 7	271 ± 25	91 ± 5
Farris	128 ± 14	102 ± 11	196 ± 14	312 ± 37	141 ± 22
Modified Farris	67 ± 8	54 ± 6	103 ± 7	190 ± 18	68 ± 6
			60 loci		
UPGMA	59 ± 3	47 ± 2	83 ± 6	143 ± 11	52 ± 4
Farris	100 ± 10	85 ± 7	146 ± 13	223 ± 18	89 ± 8
Modified Farris	54 ± 4	42 ± 3	76 ± 6	125 ± 11	48 ± 5
			100 loci		
UPGMA	47 ± 4	38 ± 3	63 ± 5	115 ± 7	40 ± 3
Farris	66 ± 6	53 ± 4	104 ± 11	158 ± 18	60 ± 8
Modified Farris	37 ± 3	29 ± 3	58 ± 5	86 ± 8	32 ± 3

method. In the present case the absolute values of genetic distances are small, so that they are approximately linear with the expected number of gene substitution in all distance measures. Therefore, overestimation of branch lengths is expected to increase the value of $\bar{S}_E$. It is also noted that UPGMA gives a somewhat smaller value of $\bar{S}_E$ than the modified Farris method. Therefore, we can conclude that with the criterion of $\bar{S}_E$ UPGMA shows the best performance.

The values of $\bar{S}_0$ for $M = 0.004$ are presented in Table 12. It is seen that the relative values of $\bar{S}_0$ for the tree-making methods and the effect of number of loci are nearly the same as those for the case of $M = 0.1$ in all distance measures. Furthermore, Tateno (1982) obtained a similar result for the case of $M = 0.002$ by using D. Therefore, whether M is large or small, the modified Farris method seems to show the best performance in estimating the branch lengths of a gene-tree, whereas the Farris method shows the poorest performance.

Discussion

We have shown that both the topology and branch length of a reconstructed tree are often quite wrong unless a large number of loci is used. In the study of phylogenetic relationships of related species in terms of electrophoresis many authors have used 20 to 40 genetic loci. Our study indicates that even if 30 loci are used and M is as large as 0.1 some parts of a reconstructed tree are incorrect with a high probability. In this study we used only 8 OTUs because of limited computer time available, but the error in reconstructed trees is expected to increase disproportionately as the number of OTUs increases (Tateno et al. 1982). One important factor for determining the accuracy of a reconstructed tree is the branch lengths of the true tree. If there are many branches of which the true distances are as small as 0.004, the reconstructed tree is usually incorrect even if 100 loci are used. This result is discouraging, but we must accept it since it is due to the stochastic nature of gene substitution. Clearly, we cannot be overconfident about the tree reconstructed from electrophoretic data. Nevertheless, our study suggests that a large part of the topology of a reconstructed tree is correct if 30 or more loci are used. In many cases even this approximate phylogenetic tree is useful for studying various evolutionary problems. Furthermore, in the future the number of loci used is expected to increase rapidly because of the recent introduction of the restriction enzyme technique in evolutionary studies.

In actual data we will almost never be sure about the reconstructed tree, unless there is some other supporting evidence. However, a rough idea about the accuracy of the tree may be obtained by the product (A) of the smallest branch length ($\hat{M}$) of the species-tree reconstructed in terms of Nei's distance and the number of loci used (r). This product ($A = r\hat{M}$) represents the total number of gene substitutions for the set of loci used, and if this value is equal to or larger than 3, the reconstructed tree seems to be quite reliable. Note that A corresponds to rM in our simulation (see Figure 1).

The accuracy of a reconstructed tree also depends on the tree-making method as well as on the distance measure used. We have seen that in estimating the branch lengths of species-trees UPGMA is the best among the three tree-making methods examined. UPGMA also seems to be the best in constructing a good topology, though when $M = 0.1$ the modified Farris method showed a slightly higher value of P than UPGMA. When the number of loci used was 70 or less, UPGMA generally showed the smallest value of d_T for both $M = 0.1$ and $M = 0.004$. It is interesting to see that in constructing a species-tree the simple UPGMA, which was originally proposed for phenetic taxonomy, shows the best performance. The reason for this seems to be that the

genetic distance based on a relatively small number of loci is subject to a large stochastic error and the procedure of distance-averaging used in UPGMA reduces this error to a considerable extent (see below).

This conclusion is somewhat different from that of our previous study (Tateno et al. 1982), where the trees based on nucleotide sequences were examined. In this case the modified Farris method showed a slightly better performance than UPGMA in topology construction except under certain circumstances. This difference apparently occured because the genetic distance based on gene frequency data is subject more often to "backward" and "parallel" changes than the estimate of nucleotide substitutions.

Note that in the presence of stochastic errors the genetic distance between a pair of populations may occasionally decrease in the evolutionary process (see below), though the expected value always increases. Estimates of nucleotide substitutions are also subject to backward and parallel mutations, but the chance of occurrency of these events is much lower in this case than in the case of gene frequency data. In our previous paper we concluded that the performance of UPGMA is better than the Farris and modified Farris methods when the coefficient of variation of branch lengths is small. However, we now realize that in the previous study the smaller coefficient of variation was associated with a larger number of nucleotide substitutions, and in this case there were many backward and parallel mutations. Therefore, we can conclude that in both cases of nucleotide substitution and gene frequency data UPGMA shows a good performance when there are many backward and parallel changes.

The fact that genetic distance occasionally decreases in the evolutionary process can be seen from Table 13 and Fig. 5. Table 13 shows one example of allele frequency changes at a locus in a simulated population for the case of M = 0.004. These frequency changes are caused by random genetic drift and mutation. It is clear that the allele frequencies for time T = 1000 x vt = 16 are considerably different from those for T = 8 but identical with those for time T = 0. Therefore, if we compute Nei's distance D from this locus alone, it becomes 0.100 between T = 0 and T = 8 but 0 between T = 0 and T = 16. Thus, the genetic distance first increases with evolutionary time and then declines to 0. Of course, computation of D from one locus is meaningless; D has to be computed from many loci. However, even if 20 to 30 loci are used, D still may decline occasionally. This can be seen from Figure 5, where the genetic distances based on 20 loci between generation 0 and other generations are presented. It is seen that D is subject to a considerable amount of random fluctuation and occasionally declines. Particularly from T = 40 to T = 48, D shows a drastic decrease. Random fluctuation of D around E(D) = 2vt decreases as the number of loci increases, but our simulation has shown that even with 100 loci the fluctuation is not negligible. We note that in UPGMA the effect of this random fluctuation is reduced because the

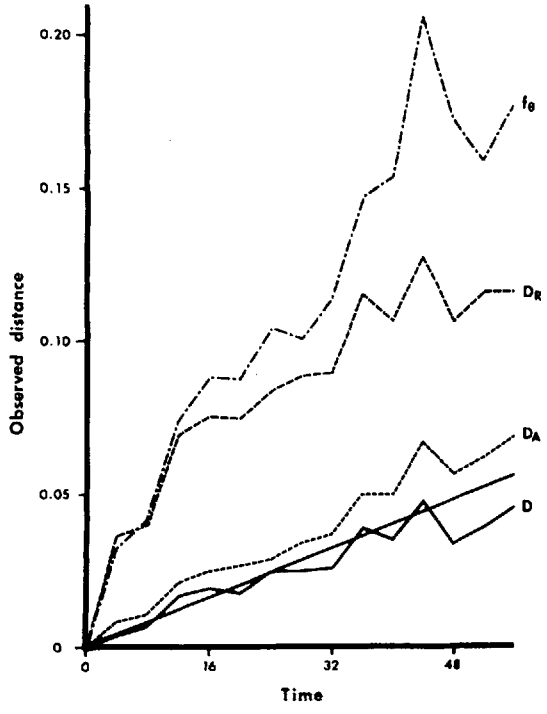

Fig. 5. Relationships between genetic distances and evolutionary time in one replication of computer simulation. Observed distance represents the distance between generation 0 and its descendant generation. 4Nv = 0.2, M = 0.004, and the number of loci used is 20. The straight line represents the expected value of D

Table 13. An example of gene frequency changes in a simulated population. Time is expressed in terms of T = 1000 × vt, where v and t are the mutation rate and generation, respectively. 4Nv = 0.2 and M = 0.004

Allele	Time							
	0	8	16	24	32	40	48	56
A_1	0.52	0.76	0.52	0.40	0.30	0.24	0.22	0.14
A_2	0.48	0.24	0.48	0.60	0.68	0.70	0.76	0.78
A_3	–	–	–	–	0.02	0.06	–	–
A_4	–	–	–	–	–	–	0.02	0.04
A_5	–	–	–	–	–	–	–	0.02
A_6	–	–	–	–	–	–	–	0.02

branch lengths of a tree are estimated by averaging several distance values. In the Farris and modified Farris methods random fluctuation of distances is assumed to represent the real genetic differences between OTUs. Figure 5 shows that D_R, f_θ, and D_A behave in essentially the same way as that of D, though they are nonlinearly related to evolutionary time. (The behavior of D_m is also essentially the same as that of D, though it is not shown here.)

We have seen that in all the three tree-making methods examined distance measure D_A tends to give a better topology than the other measures, whereas D generally gives a better estimate of branch lengths (small values of S_E) compared with the others. This suggests that D_A be used for making topology and D be used for estimating branch lengths. In this case the topology can be constructed either by the modified Farris method or by UPGMA, and once the topology is obtained, the branch lengths can be estimated by the distance-averaging method similar to that of UPGMA (Tateno et al. 1982). Of course, the advantage of D_A over D in topology construction is not always great. Furthermore, actual gene frequencies are also affected by other factors such as experimental errors, fluctuation of population size, etc., so that genetic distances are expected to be subject to even larger random errors. Therefore, we are not sure that it is always worth spending more time for this combined method rather than using UPGMA with D.

Swofford (1981) compared the efficiencies of UPGMA, Farris' (1972) method (slightly modified), and Fitch and Margoliash's (1967) methods of tree-making by using immunological or electrophoretic distance data. He concluded that the Farris method is the best and UPGMA is the poorest. This conclusion is based on the comparison of patristic distances with observed distances similar to our S_O. As shown by Tateno et al. (1982), however, S_O is not a good measure of the accuracy of a reconstructed tree. Particularly when one is interested in making a species-tree, it is a poor measure. Furthermore, even in obtaining a small value of S_O the Farris method is not necessarily better than UPGMA, as shown in the present paper.

Some numerical taxonomists claim that the genetic distance measures used in phylogeny construction should satisfy the triangle inequaltiy. They give two arguments for this. First, when one wants to represent the species of populations concerned in a multidimensional space and measure the geometric distances between them, it is necessary to use a metric that obeys this principle. Second, if every estimate of genetic distance between OTUs represents the sum of actual number of gene substitutions for all relevant branches of the true tree, then the triangle inequality should hold. Representation of populations in a multidimensional space is mathematically interesting, but it is not necessary for tree-making. Furthermore, the *geometric* distance between populations measured in this way is not proportional to the number of gene substitutions, and thus it is inappropriate for measuring *genetic* distance (Nei 1978a). (Genetic distance is the extent of genetic difference between two populations as measured by codon or nucleotide differences.) Namely, their first argument cannot be accepted.

Their second argument looks reasonable at first sight, but actually it is not realistic. As mentioned earlier, D is an asymptotic metric, and thus if a very large number of loci are used, the genetic distance between any pair of OTUs will represent the sum of the number of gene substitutions for all relevant branches at least theoretically. In practice, it is virtually impossible to examine hundreds or thousands of loci for phylogeny construction at the present time. Therefore, we must estimate the genetic distance from a smaller number of loci, and in the process of this estimation the metricity of D is disturbed by statistical errors. In other words, Nei's (1972) D is an estimate of E(D) in (2) and may deviate from this substantially when the number of loci is small. Nevertheless, it is possible to construct a reasonable phylogenetic tree by using D, as shown in this paper. Actually metricity is not really required for tree-making as long as a proper distance measure and a proper tree-making method are used. In this connenction it should be noted that usual estimates of nucleotide or amino acid substitutions are also a nonmetric, because they are estimated statistically by taking into account back mutations, parallel mutations, and multiple mutations, and essentially the same argument as the above applies to these estimates (see Tateno et al. 1982).

Recently, Farris (1981) criticized Sarich and Wilson's (1967) immunological distance and Nei's (1972) distance for their nonmetricity, and claimed that any nonmetric distance cannot show clocklike behavior. He apparently failed to see that molecular clock is stochastic rather than deterministic and subject to errors due to backward and parallel mutations. Nonmetricity itself does not deprive clocklike behavior. Furthermore, his criticism on D is based on gene frequency differences for one locus. In practice, D is designed to be used for many loci and should never be used for one locus (Nei 1972). Note also that, unlike Farris' assumption, the amount of gene frequency difference between two populations is not proportional to evolutionary time whether there is selection or not. This is an elementary knowledge in population genetics. In Farris' view no negative estimates of branch lengths are allowed in tree construction. However, we have shown that negative branch length can occur legitimately when distance estimates are based on a relatively small number of loci (Figure 5). They are caused by sampling error not by the defect of the distance measure used. If we forbid negative estimates, the estimates will no longer be unbiased, and this biasedness will introduce another source of errors in tree-making. Considering gene frequency differences at a single locus, Farris (1981) stated that none of the

distance measures currently available permits path length interpretations. As mentioned above, however, genetic distance should not be computed from a single locus. It is also very important to know the dynamics of gene frequency changes in populations for making a proper interpretation of branch lengths. Only when this dynamics is taken into account properly, can one develop a distance measure that is proportional to evolutionary time. D has been developed exactly in this way.

As mentioned earlier, Rogers' (1972) distance (D_R) has often been used in conjunction with the Farris method, because it satisfies the triangle inequality. The fact that this distance does not give negative branches when the Farris method is used seems to have been attractive to some workers. However, metricity of distance itself does not give any advantage in tree-making, as we have seen in our computer simulation. Just like D, this distance may occasionally decrease in the evolutionary process because of stochastic errors, and thus D_R does not necessarily show the true genetic relationship among OTUs. For example, in Table 3 the observed values of D_R for OTUs 1 and 6, 1 and 7, and 6 and 7 are 0.678, 0.637, and 0.733, respectively. These values suggest that OTUs 1 and 7 are most closely related among the three OTUs. In the true tree (Fig. 1a), however, 1 and 6 are more closely related than 1 and 7. In the Farris method observed values of genetic distances are assumed to represent *true distances*, but as is clear from this example they are generally quite different from true distances. This erroneous assumption is also one of the reasons for the poor performance of this method. In UPGMA observed distances are regarded as *estimates of true distances*, and with this understanding branch lengths of a reconstructed tree are estimated (cf Nei 1975). Incidentally, UPGMA also does not produce negative estimates of branch lengths, and if the topology reconstructed is correct, UPGMA gives least-squares estimates of branch lengths (Chakraborty 1977).

At this point it should be indicated that in addition to its nonlinear relationship with evolutionary time D_R has another theoretical defect. Namely, it is not necessarily 1 even when the two populations concerned have no shared alleles. This occurs when the populations are polymorphic. For example, when there are five nonshared alleles in each population and all allele frequencies are equal, i.e. 1/5, we have $D_R = 1/\sqrt{5} = 0.45$.On the other hand, if the two populations are fixed for different alleles, D_R becomes 1. From the genetic point of view, this is a poor property. A similar property is observed with D_m or the Manhattan distance ($\Sigma\Sigma|x_{ij} - y_{ij}|$) which is often advocated by numerical taxonomists.

In the present study we have assumed no selection. Our conclusion, however, would not change very much in the presence of advantageous mutation, as long as the expected rate of gene substitution remains constant. This is because the distribution of D is not affected appreciable by advantageous selection (Chakraborty et al. 1977). However, our assumption of constant rate of gene substitution may not hold under certain circumstances. This is particularly so with gene frequency data, because the extent of genetic distance is affected by the bottleneck effect (Chakraborty and Nei 1977). When long-term evolution is considered, the effect of varying rate of gene substitution is generally to increase the variance of genetic distance, and thus it increases the relative merit of UPGMA over the other methods (Tateno et al. 1982). When short-term evolution is considered and the bottleneck effect on genetic distances remains or when the entire set of genes is subjected to the same directional change for a relatively short period of time, the modified Farris method is expected to be better than UPGMA. However, a more detailed study should be conducted on this point.

In our simulation we used 8 OTUs to save computer time. However, Tateno et al.'s (1982) study with nucleotide sequences for 32 OTUs suggests that as the number of OTUs increases, the distortion index (d_T) of a reconstructed tree rapidly increases and when the number is large the chance of obtaining the correct topology is extremely small unless the number of loci used is very large. Therefore, it is advisable not to include many OTUs in tree making. If OTUs to be studied can be divided into groups, it seems better to construct a tree first among groups and then within groups.

In the present paper we have been concerned with construction of phylogenetic trees from gene frequency data. In recent years evolutionists (e.g. Brown et al. 1979; Avise et al. 1979; Shah and Langley 1979) started to use the restriction endonuclease technique to study the genetic differences between species or populations. In this case the number of nucleotide differences per nucleotide site can be estimated by the statistical methods of Nei and Li (1979), Kaplan and Langley (1979), and Gotoh et al. (1979). The estimates obtained by these methods have a statistical property similar to that of Nei's D. Therefore, the conclusions obtained in this paper seem to apply to these estimates as well. In this case, however, we must increase the number of endonucleases used to raise the accuracy of a reconstructed tree.

Acknowledgements. We thank Takashi Gojobori for his comments on the manuscript. This study was supported by research grants from the National Science Foundation and the National Institutes of Health.

References

Avise JC, Lansman RA, Shade RO (1979) The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. I. Population structure and evolution in the genus Peromyscus. Genetics 92:279–295

Bhattacharrya A (1946) On a measure of divergence between two multinomial populations. Sankhya 7:401–406

Brown WM, George Jr. M, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci 76:1967–1971

Cavalli-Sforza LL (1969) Human Diversity. Proc 12th Intl Cong Genet, Tokyo, Vol 3:405–416

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

Cavalli-Sforza LL, Piazza A (1975) Analysis of evolution: Evolutionary rates, independence and treeness. Theoret Pop Biol 8:127–165

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

Chakraborty R, Nei M (1977) Bottleneck effects on average heterozygosity and genetic distance with the stepwise mutation model. Evolution 31:347–356

Chakraborty R, Fuerst PA, Nei M (1977) A comparative study of genetic variation within and between populations under the neutral mutation hypothesis and the model of sequentially advantageous mutations. (Abstract) Genetics 86:s10–11

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

Farris JS (1981) Distance data in phylogenetic analysis. In: Funk VA, Brooks DR (eds) Advances in cladistics. Proc. 1st Meeting of Willi Hennig Society, Publ. New York Botanical Garden, Bronx, NY, pp 1–23

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

Gotoh O, Hayashi JI, Yonekawa H, Tagashira Y (1979) An improved method for estimating sequence divergence between related DNAs from changes in restriction endonuclease cleavage sites. J Mol Evol 14:301–310

Griffiths RC (1980) Lines of descent in the diffusion approximation of neutral Wright-Fisher models. Theoret Pop Biol 17:37–50

Griffiths RC, Li WH (1983) Simulating allele frequencies in a population and the genetic differentiation of populations under mutation pressure. Theoret Pop Biol (in press)

Kaplan N, Langley CH (1979) A new estimate of sequence divergence of mitochondrial DNA using restriction endonuclease mappings. J Mol Evol 13:295–304

Kidd KK, Cavalli-Sforza LL (1971) Number of characters examined and error in reconstruction of evolutionary trees. In: Hodson FR, Kendall DG, Tautu P (eds) Mathematics in the archaeological and historical sciences. Edinburgh University Press, Edinburgh, pp 335–346

Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. Genetics 49:725–738

Li WH (1976) Effect of migration on genetic distance. Amer Nat 110:841–847

Li WH, Nei M (1975) Drift variances of heterozygosity and genetic distance in transient states. Genet Res 25:229–248

Nei M (1972) Genetic distance between populations. Amer Nat 106:283–292

Nei M (1973) The theory and estimation of genetic distance. In: Morton NE (ed) Genetic structure of populations. University of Hawaii Press, Honolulu, pp 45–54

Nei M (1975) Molecular population genetics and evolution. North Holland, Amsterdam and New York

Nei M (1976) Mathematical models of speciation and genetic distance. In: Karlin S, Nevo E (eds) Population genetics and ecology. Academic Press, New York, pp 723–765

Nei M (1977) Standard error of immunological dating of evolutionary time. J Mol Evol 9:203–211

Nei M (1978a) The theory of genetic distance and evolution of human races. Japan J Hum Genet 23:341–369

Nei M (1978b) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583–590

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci 76:5269–5273

Nei M, Roychoudhury AK (1974) Sampling variances of heterozygosity and genetic distance. Genetics 76:379–390

Nei M, Tateno Y (1975) Interlocus variation of genetic distance and the neutral mutation theory. Proc Natl Acad Sci 72: 2758–2760

Prager EM, Wilson AC (1978) Construction of phylogenetic trees for proteins and nucleic acids: comparison of alternative matrix methods. J Mol Evol 11:129–142

Robinson DF, Foulds LR (1981) Comparison of phylogenetic trees. Math Biosci 53:131–147

Rogers JS (1972) Measures of genetic similarity and genetic distance. Studies in Genetics VII (University of Texas Publ. No. 7213), pp 145–153

Sanghvi LD (1953) Comparison of genetical and morphological methods for a study of biological differences. Amer J Phys Anthrop 11:385–404

Sarich VM, Wilson AC (1967) Immunological time scale for hominid evolution. Science 158:1200–1203

Shah DM, Langley CH (1979) Inter- and intraspecific variation in restriction maps of *Drosophila* mitochondrial DNAs. Nature 281:696–699

Sneath PHA, Sokal RR (1973) Numerical taxonomy. WH Freeman, San Francisco

Swofford DL (1981) On the utility of the distance Wagner procedure. In: Funk VA, Brooks DR (eds) Advances in cladistics. Proc. 1st Meeting of Willi Hennig Society, Publ. New York Botanical Garden, Bronx, NY, pp 25–43

Tateno Y (1982) Statistical examination of phylogenetic tree construction methods by computer simulation. In: Kimura M (ed) Molecular evolution, protein polymorphism and the neutral theory. Japan Scientific Societies Press, Tokyo/ Springer-Verlag, Berlin, pp 217–229

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 20/Revised November 20, 1982