AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 84:75–85 (1991)

Reconstruction of Molecular Phylogeny of Extant Hominoids From DNA Sequence Data

NARUYA SAITOU
Department of Anthropology, Faculty of Science, The University of Tokyo, Tokyo 113, Japan.

KEY WORDS Human-chimpanzee-gorilla triad, Nucleotide substitutions, Neighbor-joining method

ABSTRACT Evolutionary distance matrices of the extant hominoids are computed from DNA sequence data, and hominoid DNA phylogenies are reconstructed by applying the neighbor-joining method to these distance matrices. The chimpanzee is clustered with the human in most of the phylogenetic trees thus obtained. The proportion of the distance between human and chimpanzee to that between human/chimpanzee and orangutan is estimated. Both mitochondrial DNA and nuclear DNA show a similar value (0.44), which is close to values derived from DNA-DNA hybridization data.

The phylogeny of extant hominoids has been studied extensively by a variety of molecular techniques such as microcomplement fixation (Sarich and Wilson, 1967), amino acid sequences (Goodman et al., 1971), DNA-DNA hybridization (Kohne et al., 1972), and restriction enzymes (Ferris et al., 1981). Recently, several portions of mitochondrial and nuclear DNA of hominoids have been sequenced (e.g., Brown et al., 1982; Koop et al., 1986). Despite the large data set that has been accumulated, however, there is no consensus on the problem of the branching order of human, chimpanzee, and gorilla. The divergence time among these three species is also in dispute.

A point that complicates the determination of phylogenetic branching patterns is that different tree-making methods can produce different trees even if the same data set is used. For example, chimpanzee and gorilla are clustered as sister taxa when the maximum parsimony method (Eck and Dayhoff, 1966; Fitch, 1977) is applied to mitochondrial DNA sequence data (Brown et al., 1982), but human and chimpanzee are clustered by applying UPGMA (Sokal and Sneath 1963) to the same data set (Nei et al., 1985). Thus the choice of the tree-making method should also be considered.

Saitou and Nei (1987) proposed the neighbor-joining method for phylogenetic tree reconstruction from evolutionary distance matrices. The neighbor-joining method does not depend on the assumption of a molecular clock and has been shown to find the correct unrooted tree efficiently (Saitou and Nei, 1987; Sourdis and Nei, 1988; Saitou and Imanishi, 1989).

In the present study, we first estimate the evolutionary distances from a variety of DNA sequence data and apply the neighbor-joining method to these distances for reconstructing the molecular phylogenies. Then we compare the evolutionary distances for the trio of human, chimpanzee, and orangutan and show that the proportions of the distance between human and chimpanzee to that between human/chimpanzee and orangutan are remarkably similar for both mitochondrial and nuclear DNA data.

NUCLEOTIDE SEQUENCES USED

We used nucleotide sequence data for two fragments of the mitochondrial DNA and for four fragments of the nuclear DNA. The sequence data includes human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), pygmy chimpanzee (*Pan paniscus*), gorilla (*Gorilla gorilla*), orangutan (*Pongo pygmaeus*), gibbon (*Hylobates lar*), rhesus monkey (*Macaca mulatta*), squirrel monkey (*Saimiri sciureus*), owl monkey (*Aotus trivirgatus*), and spider monkey (*Ateles geoffroyi*). Nucleotide sequence data used are as follows.

Received October 11, 1989; accepted April 24, 1990.

1. A 0.9-kb fragment of mitochondrial DNA containing genes for NADH dehydrogenase subunits 3 and 4 and for three tRNAs. Nucleotide sequences for human, chimpanzee (a part of the sequence is that of pygmy chimpanzee), gorilla, orangutan, and gibbon are taken from Brown et al. (1982); those for rhesus monkey and squirrel monkey and the alignment of these seven sequences are from Hayasaka et al. (1988). This DNA fragment is designated as mtDNA I.

2. A 0.9-kb fragment containing the small ribosomal RNA gene of mitochondrial DNA. The human nucleotide sequence is from Anderson et al. (1981), and those for chimpanzee, pygmy chimpanzee, gorilla, and orangutan are from Hixson and Brown (1986). The alignment for these five sequences follows that of Hixson and Brown (1986). This DNA fragment is designated as mtDNA II.

3. A 2.0-kb fragment containing the η-globin pseudogene in the β-globin gene family. Sequences for human, chimpanzee, and gorilla are taken from Chang and Slightom (1984); that for owl monkey is from Harris et al. (1984); and those for orangutan and rhesus monkey are from Koop et al. (1986). The alignment for these seven sequences is from Koop et al. (1986). This DNA fragment is designated as η-globin S.

4. A 3.1-kb fragment containing a spacer DNA of the β-globin gene family. Nucleotide sequence for two alleles (R and T) of human and that for chimpanzee are taken from Maeda et al. (1983), and those for gorilla, orangutan, rhesus monkey, and spider monkey as well as the alignment of all seven sequences are from Maeda et al. (1988). This DNA fragment is designated as β-globin.

5. A 7.1-kb fragment containing the η-globin pseudogene that includes the η-globin S fragment. The human nucleotide sequence is taken from Collins and Weissman (1984), and those for chimpanzee, gorilla, and orangutan (excluding the part of η-globin S) as well as the alignment of all four sequences are from Miyamoto et al. (1987). This DNA fragment is designated as η-globin L.

6. A 2.3-kb fragment containing immunoglobulin $\epsilon 3$ processed pseudogene. The human nucleotide sequence is from Ueda et al. (1982), and those for chimpanzee, gorilla, and orangutan are from Ueda et al. (1989). This DNA fragment is designated as Ig-$\epsilon 3$.

EVOLUTIONARY DISTANCE MATRICES

There are a variety of methods for estimating evolutionary distances based on the accumulation of nucleotide substitutions (e.g., Jukes and Cantor 1969; Kimura, 1980; Gojobori et al., 1982). Tajima and Nei (1984) proposed a simple method that gives a good estimate of evolutionary distance compared with the other methods. This method incorporates the nonrandom pattern of nucleotide substitution that has been observed in nuclear genes (Li et al., 1984) and in mitochondrial genes (Aquadro et al., 1984). Thus we used Tajima and Nei's (1984) equation 6 for estimating the evolutionary distance (d) from the proportion (p) of different nucleotides per site, the proportion (q_i) of the ith nucleotide (i = A, G, T, or C), and the proportion (x_{ij}) of a pair of nucleotides i and j between the two sequences compared:

$$d = -b \log (1 - p/b), \qquad (1)$$

where

$$b = (1 - \sum_{i=1}^{4} q_i^2 + p^2/h)/2$$

and

$$h = \sum_{i=1}^{3} \sum_{j=i+1}^{4} x_{ij}^2/(2q_iq_j).$$

The standard error of d is given by the square root of the variance of d,

$$V(d) = b^2p(1 - p)/[(b - p)^2n], \qquad (2)$$

where n is the number of nucleotides compared (Tajima and Nei, 1984). Equation 1 resembles Jukes and Cantor's (1969) formula in which the random nucleotide substitution is assumed. In fact, equation 1 is reduced to Jukes and Cantor's formula when b = 0.75. When the pattern of nucleotide substitution is not random, however, the value of b (estimated by using observable quantities q_is and x_{ij}s) becomes different from 0.75, and this is why equation 1 is efficient in obtaining a reliable estimates of nucleotide substitutions.

Evolutionary distance matrices thus obtained are presented in Table 1. Gaps (insertions and deletions) were excluded from the comparison.

Table 1a shows the distance matrix for the mtDNA I fragment. Nei et al. (1985) presented a similar distance matrix for five hominoid species by applying Jukes and Cantor's (1969) method, whereas Hayasaka

TABLE 1. Evolutionary distance (SE) matrices

a. mtDNA I (892 nucleotides were compared)

	Human	Chimpanzee	Gorilla	Orangutan	Gibbon	Rhesus monkey
Chimpanzee	0.0976 (0.0116)					
Gorilla	0.1141 (0.0127)	0.1183 (0.0130)				
Orangutan	0.1880 (0.0173)	0.2046 (0.0184)	0.1966 (0.0179)			
Gibbon	0.2157 (0.0188)	0.2280 (0.0196)	0.2278 (0.0196)	0.2264 (0.0194)		
Rhesus monkey	0.2929 (0.0230)	0.3234 (0.0251)	0.2939 (0.0231)	0.3150 (0.0245)	0.2962 (0.0231)	
Squirrel monkey	0.3644 (0.0269)	0.3802 (0.0280)	0.3543 (0.0263)	0.3689 (0.0268)	0.3473 (0.0255)	0.3967 (0.0290)

b. mtDNA II (939 nucleotides were compared)

	Chimpanzee	Pygmy chimpanzee	Gorilla	Human
Pygmy chimpanzee	0.0118 (0.0036)			
Gorilla	0.0428 (0.0069)	0.0416 (0.0068)		
Human	0.0383 (0.0065)	0.0327 (0.0060)	0.0371 (0.0064)	
Orangutan	0.0954 (0.0106)	0.0917 (0.0104)	0.0966 (0.0107)	0.0929 (0.0105)

c. η-globin S (1,967 nucleotides were compared)

	Human	Chimpanzee	Gorilla	Orangutan	Rhesus monkey
Chimpanzee	0.0123 (0.0025)				
Gorilla	0.0144 (0.0027)	0.0181 (0.0031)			
Orangutan	0.0308 (0.0041)	0.0346 (0.0043)	0.0373 (0.0045)		
Rhesus monkey	0.0734 (0.0065)	0.0780 (0.0067)	0.0781 (0.0067)	0.0785 (0.0067)	
Owl monkey	0.1133 (0.0083)	0.1145 (0.0083)	0.1170 (0.0085)	0.1174 (0.0085)	0.1369 (0.0093)

d. β-globin (3,042 nucleotides were compared)

	Human R	Human T	Chimpanzee	Gorilla	Orangutan	Rhesus monkey
Human T	0.0053 (0.0013)					
Chimpanzee	0.0133 (0.0021)	0.0140 (0.0022)				
Gorilla	0.0150 (0.0023)	0.0157 (0.0023)	0.0123 (0.0020)			
Orangutan	0.0300 (0.0032)	0.0307 (0.0033)	0.0255 (0.0030)	0.0269 (0.0030)		
Rhesus monkey	0.0778 (0.0054)	0.0774 (0.0054)	0.0742 (0.0052)	0.0731 (0.0052)	0.0701 (0.0051)	
Spider monkey	0.1172 (0.0068)	0.1169 (0.0068)	0.1118 (0.0066)	0.1135 (0.0067)	0.1138 (0.0067)	0.1331 (0.0073)

e. η-globin L (6,902 nucleotides were compared)

	Human	Chimpanzee	Gorilla
Chimpanzee	0.0144 (0.0015)		
Gorilla	0.0145 (0.0015)	0.0183 (0.0017)	
Orangutan	0.0297 (0.0021)	0.0338 (0.0023)	0.0334 (0.0023)

TABLE 1. Evolutionary distance (SE) matrices (continued)

f. Ig-ϵ3 (2,250 nucleotides were compared)

	Human	Chimpanzee	Gorilla
Chimpanzee	0.0139 (0.0025)		
Gorilla	0.0176 (0.0028)	0.0162 (0.0027)	
Orangutan	0.0409 (0.0044)	0.0385 (0.0043)	0.0414 (0.0044)

et al. (1988) used Gojobori et al.'s (1982) method to estimate the evolutionary distances. All distance values of Table 1a are slightly larger than those of these two distance matrices. Saitou (1988) computed the evolutionary distances for the data of the mtDNA II fragment by applying Jukes and Cantor's (1969) method, and the distance values of the present study, shown in Table 1b, are slightly larger than those. Maeda et al. (1988) and Ueda et al. (1988) used Kimura's (1980) method for estimating the evolutionary distances for the data of the β-globin fragment and those of the Ig-ϵ3 fragment, respectively. These distances are similar to those presented in Table 1d and Table 1f, respectively.

PHYLOGENETIC TREES

Phylogenetic trees were reconstructed by applying the neighbor-joining method to the distance matrices of Table 1. The principle of the neighbor-joining method is to find pairs of operational taxonomic units (OTUs) that minimize the total branch length at each stage of clustering OTUs, starting with a star-like tree. The branch lengths as well as the tree topology of a minimum-evolution tree can quickly be obtained by applying this stepwise clustering method. In fact, it has been shown that the efficiency of the neighbor-joining method in obtaining the correct tree topology is similar to that of the minimum-evolution method (Saitou and Imanishi, 1989). The neighbor-joining method gives an unrooted tree (as do many other tree-making methods), and a New World monkey (either squirrel, owl, or spider monkey) was accordingly taken as an outgroup species and used to locate the root of each tree.

Figure 1a shows the phylogeny of the mtDNA I fragment for human, chimpanzee, gorilla, orangutan, gibbon, and rhesus monkey. Squirrel monkey was used as the outgroup to locate the root. Human and chimpanzee are clustered in the tree of Figure 1a. The branching pattern of this tree is the same as that of Nei et al. (1985), in which UPGMA was applied to the evolutionary distances estimated by Jukes and Cantor's (1969) method. Nei et al. showed that the branching point for human and chimpanzee lineage was not significantly different from

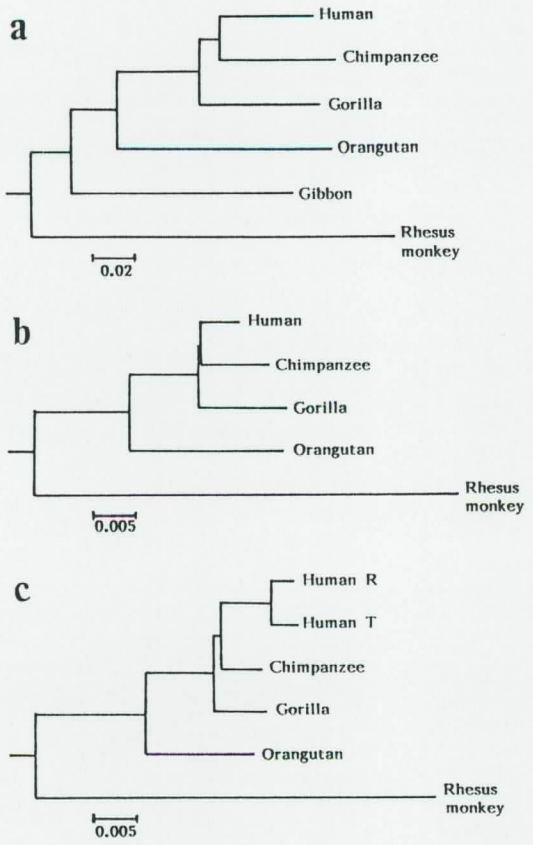

Fig. 1. Phylogenetic trees of higher primates reconstructed by the neighbor-joining method from the evolutionary distance matrices of Table 1. a: mtDNA I. b: η-Globin S. c: β-Globin. Scales are the number of nucleotide substitutions per nucleotide site.

that for the human-chimpanzee lineage and the gorilla lineage in their tree.

This tree was also chosen when Fitch and Margoliash's (1967) method and the distance Wagner method (Farris, 1972) were used (Nei, 1987). Hasegawa et al. (1985) obtained the same tree by applying Felsenstein's (1981) maximum likelihood method to the transversional difference alone. Using a slightly different maximum likelihood method, Saitou (1988) obtained a tree in which chimpanzee and gorilla were clustered. The same tree was obtained by using the maximum parsimony method (Brown et al., 1982). Interestingly, however, we obtain an alternative tree (in which human and chimpanzee are clustered) by the maximum parsimony method if sequences for only human, chimpanzee, gorilla, and orangutan are used (Saitou and Nei, 1986). The same tree is obtained by Lake's (1987) evolutionary parsimony method (Holmquist et al., 1988).

The rhesus monkey seems to have a higher rate of nucleotide substitutions than that of hominoids, since the branch length for the rhesus monkey lineage is longer than those for any of the hominoid lineages. Hayasaka et al. (1988) also applied the neighbor-joining method to the evolutionary distances estimated by using Gojobori et al.'s (1982) method and obtained a tree in which human and chimpanzee were clustered. Their tree included sequences for three other Old World monkeys (Japanese macaque, crab-eating macaque, and Barbary macaque) in addition to those for two prosimians (tarsier and lemur). The inclusion of such distantly related species probably diluted the rate difference between the hominoid lineage and the Old World monkey lineage in their tree. However, a more detailed comparison revealed a higher rate of nucleotide substitution in the Old World monkey lineage than that in the hominoid lineage (Hayasaka et al., 1988).

When the data for mtDNA II were used, human and gorilla were clustered (tree not shown). However, the length of the branch connecting human-gorilla cluster with chimpanzee cluster is very short. The branching pattern and branch lengths of this tree are quite similar to those of Saitou (1988), in which the neighbor-joining method was used to Jukes-Cantor distances [see Saitou (1991) and Saitou and Imanishi (1989) for results when the other tree-making methods were used to the same sequence data]. The maximum likelihood tree of Saitou (1988) is also the same tree. On the other hand, this tree and the one that clusters human and chimpanzee are equally parsimonious if we apply the maximum parsimony method (Hixson and Brown, 1986). It is clear from the distance matrix of Table 1b that chimpanzee and pygmy chimpanzee are clustered followed by the clustering of human and the chimpanzee-pygmy chimpanzee cluster, if we apply UPGMA.

Figure 1b is a nuclear DNA phylogeny of the η-globin S fragment for four hominoid species as well as an Old World monkey species (rhesus monkey). The owl monkey was used to locate the root. Human and chimpanzee are clustered in this tree, but the branch length from the human-chimpanzee cluster to the gorilla lineage is very short. The branching pattern of this tree is the same as that obtained by Li's (1981) method (Li and Tanimura, 1987) and by the maximum likelihood method (Hasegawa et al., 1987). This tree and the one in which chimpanzee and gorilla were clustered were equally parsimonious (Koop et al., 1986). If we apply UPGMA to the distance matrix of Table 1c, human and chimpanzee are clustered.

Figure 1c is a nuclear DNA phylogeny of the β-globin fragment for the same set of primate species as in the case of Figure 1b. Spider monkey is used as an outgroup species in this case. Human and chimpanzee are clustered in this tree as in Figure 1b. The branching pattern of this tree is the same as that obtained by the maximum parsimony method, whereas chimpanzee and gorilla are clustered first if we use UPGMA (Maeda et al., 1988), and human and gorilla are clustered first when the maximum likelihood method is used (Hasegawa et al., 1989).

The data available for both the η-globin L and Ig-ϵ3 fragments include nucleotide sequences for human, chimpanzee, gorilla, and orangutan. Human and chimpanzee are clustered if we apply the neighbor-joining method or UPGMA to the evolutionary distance matrices of Table 1e and Table 1f (trees not shown). This tree was also obtained when the maximum parsimony method was used (Miyamoto et al., 1987; Ueda et al., 1989).

COMPARISON OF EVOLUTIONARY DISTANCES

It now seems clear from the phylogenetic tree analysis that the rate of nucleotide substitutions is more or less constant in the

TABLE 2. Comparison of evolutionary distances for human, chimpanzee, and orangutan

DNA fragment	N[1]	D (h − c) = A	D (h − o) = B	D (c − o) = C	2A / B + C
mtDNA I	895	0.0973	0.1889	0.2055	0.49
mtDNA II	945	0.0386	0.0993	0.1019	0.38
Average of mitochondrial DNA					0.44
η-Globin L	6903	0.0144	0.0297	0.0338	0.45
β-Globin	3150	0.0128	0.0306	0.0263	0.45
Ig-ε1	2190	0.0181	0.0353	0.0339	0.52
Ig-ε3	2250	0.0139	0.0409	0.0385	0.35
Average of nuclear DNA					0.44

[1]Number of nucleotides compared.

hominoid lineage. We accordingly undertook a more detailed comparison of the evolutionary distances among human, chimpanzee, and orangutan. The result is shown in Table 2. In this table, data for the η-globin S fragment were not included, since this fragment is a part of the η-globin L fragment. On the other hand, data for the DNA fragment designated as Ig-ε1 are included in Table 2. This fragment contains the Ig-ε1 functional gene and nucleotide sequences for human, chimpanzee, and orangutan (Ueda et al., 1982; Sakoyama et al., 1987) are available. The Ig-ε1 and -3 fragments are in different chromosomes in human (Ueda et al., 1982). Note that the numbers of nucleotides for the other DNA fragments have been increased compared to those of Table 1, because the insertions common to human, chimpanzee, and orangutan that were excluded from the previous analysis are now included. D(i–j) denotes the evolutionary distance between species i and j, and h, c, and o designate human, chimpanzee, and orangutan, respectively.

As expected, mitochondrial DNA exhibits larger evolutionary distances than those of nuclear DNA. In particular, mtDNA I has the largest amount of divergence. On the other hand, four DNA fragments of nuclear DNA show more or less the same amount of divergence, 0.013–0.018 between human and chimpanzee and about 0.03 between human and orangutan and between chimpanzee and orangutan. It is interesting to note that apart from Ig-ε1, the fragments of nuclear DNA are nonfunctional DNAs (either a pseudogene or a spacer region). Thus these results seem to be in agreement with the prediction of the neutral theory that the mutation rate is the upper limit for the rate of nucleotide substitutions in nonfunctional DNAs (Kimura, 1983). As for the functional Ig-ε1 gene, the selection constraint against this gene may be small, because its d values were as large as that for nonfunctional DNA fragments (Table 2). Although the evolutionary distance between human and chimpanzee for this gene is slightly larger than those for nonfunctional DNAs, this may be attributable to sampling errors. Distances between human and orangutan (B) and those between chimpanzee and orangutan (C) are more or less the same. This indicates that the evolutionary rates of the human and the chimpanzee lineages are similar to each other.

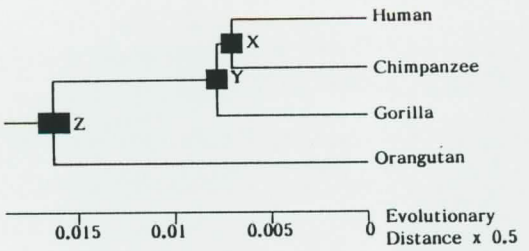

Fig. 2. A molecular phylogeny of hominoids estimated by the combined nuclear DNA sequence data. See text for details.

The rightmost column in Table 2 shows the proportion of the distance (A) between human and chimpanzee to that of the distance between human-chimpanzee cluster and orangutan. The latter distance was obtained by averaging distances B and C. Thus the values of 2A/(B + C) correspond to D(X − human)/D(Z − orangutan), where X and Y are branching points for human and chimpanzee and for the human-chimpanzee cluster and orangutan (see Fig. 2). Interestingly, all the proportions are more or less the same, and the simple averages for the mitochondrial DNA and that for the nuclear DNA both become 0.44.

NUCLEAR DNA PHYLOGENY OF HOMINOIDS

As shown in the phylogenetic analysis, all the nuclear DNA data produced the same tree for human, chimpanzee, gorilla, and orangutan, and the approximate constancy of the rate of nucleotide substitution has been observed for these four hominoid species. Thus we combined all the nuclear DNA data (η-globin L, β-globin, and Ig-ϵ3) and made the combined distance matrix (Table 3). A phylogenetic tree (Fig. 2) was constructed by applying UPGMA, since this tree-making method assumes a constant rate of evolution (see Nei, 1987). The sequence for allele R was used as the human sequence for the case of the β-globin fragment. The total number of nucleotides compared is 12,297 bp. As expected, human and chimpanzee are clustered in the tree of Figure 2, because the pair of OTUs that show that the smallest distance is clustered first in UPGMA. This clustering was also obtained when the neighbor-joining method was applied to the distance matrix of Table 3 (tree not shown).

Distances between the three branching points (X, Y, and Z) and the extant species are 0.00695, 0.00783, and 0.01630, respectively (see Fig. 2). Standard errors of these branching points (or more precisely, those of the distances from these branching points to the present time) were obtained by applying Nei et al.'s (1985) method. Since Tajima and Nei's (1984) method was used for distance estimation, equations 1 and 2 were used instead of equations 14 and 15 of Nei et al. (1985), which are the special cases of equations 1 and 2, with b = 0.75.

The standard error of the branching point X is 0.000537, which is simply one-half of the standard error of D (human-chimpanzee) (see Table 3). Computation of the standard error of the branching point Y (d_Y) is more complicated. Because d_Y is estimated by [D (gorilla-human) + D (gorilla-chimpanzee)]/4, the variance of d_Y becomes

$$V(d_Y) = [V(D_{gh}) + V(D_{gc}) + 2Cov(D_{gh}, D_{gc})]/16, \quad (3)$$

where V(x) and Cov(x, y) denotes the variance of x and covariance between x and y, respectively, and D_{xy} is D(x − y) (Nei et al., 1985). Species names were abbreviated as g, gorilla; h, human; c, chimpanzee. Cov (D_{gh}, D_{gc}) is identical to $V(D_{gx})$, and this value can be obtained by putting the *P* value corresponding to D (gorilla − x) into equation 2. We used b = 0.6 in equation 1 for converting the distance to the *P* value. Hence the variance of the branching point Y can be computed. Computation of the variance of the branching point Z can also be done in a similar manner (see Nei et al., 1985, for details).

Therefore, the standard errors (square roots of the corresponding variances) of the branching points Y and Z become 0.000503 and 0.000765, respectively. These standard errors are graphically represented in Figure 2 as boxes.

We can test whether the difference between the branching points X and Y of Figure 2 is statistically significant or not by using a t test (Nei et al., 1985). The difference between the two branching points is 0.00088 (= 0.00783 − 0.00695) and its variance is estimated by $0.000537^2/2 + 0.000503^2 = 3.97 \times 10^{-7}$. Because there are correlations between the two branching points, this variance is slightly smaller than that used for the conventional t test (see Appendix to Nei et al., 1985). In any case, the t statistic becomes 1.40 (= 0.00088/0.00063). This t value is not statistically significant ($P > 0.081$; one-tailed test with d.f. = ∞). Therefore, the clustering of human and chimpanzee in Figure 2 is not a definitive one. As for the branching points Y and Z, the difference between these two branching points is highly significant (t = 0.00847/0.000844 = 10.04; $P < 0.0001$). This indicates that the trio of human, chimpanzee, and gorilla is a well-defined monophyletic group, and that the orangutan is clearly an outgroup to the trio.

TABLE 3. Evolutionary distance (SE) among four hominoids

	Human	Chimpanzee	Gorilla
Chimpanzee	0.01390 (0.00107)		
Gorilla	0.01508 (0.00112)	0.01624 (0.00116)	
Orangutan	0.03189 (0.00165)	0.03265 (0.00167)	0.03328 (0.00169)

DISCUSSION

Branching order of human, chimpanzee, and gorilla

The determination of the branching order of human, chimpanzee, and gorilla is per-

haps the most important current issue in human phylogeny. Although human and chimpanzee are clustered in the tree of Figure 2, this clustering is not statistically significant. Using a different statistical test, Holmquist et al. (1988), Li (1989), and Holmes et al. (1989) also obtained a nonsignificant result for the nuclear DNA sequence data, although the data for Ig-ϵ3 were not used in these studies.

Recently, Williams and Goodman (1989) claimed that they obtained a significant clustering of human and chimpanzee from the same data set used in the present study based on the maximum parsimony method. Gaps (insertions and deletions) as well as nucleotide substitutions were used in their study, whereas gaps were not used in the present study. Interestingly, all four characters involving gaps used in Williams and Goodman's study support the clustering of human and chimpanzee. Thus the significance of their result may disappear if these four characters are excluded from the comparison. Because gaps are introduced when nucleotide sequences are aligned, their existence or nonexistence is somewhat artificial. Furthermore, the nature of gap creation is not clear compared with the well documented nature of nucleotide substitutions (e.g., Nei, 1987). It should also be noted that different phylogenetic tree-making methods can yield different branching patterns for the same data, as described in a previous section. This indicates that the available amount of sequence data is not enough to determine the branching order of the trio. Therefore, it may be said that the branching order of human, chimpanzee, and gorilla is not yet resolved at the DNA sequence level.

There is another important aspect for the determination of the branching pattern of these three hominoid species. In the presence of genetic polymorphism, a phylogenetic tree reconstructed from a single gene from each species (gene tree) may be different from the true phylogenetic tree of species (species tree), even if a large number of nucleotides are used (Takahata and Nei, 1985; Saitou and Nei, 1986; Nei, 1987). This is especially so when the two speciation events occurred in a short period. Because there is a tight linkage between the mtDNA I and mtDNA II fragments and between the η-globin and β-globin fragments, we can obtain three different gene trees for mitochondria, the β-globin gene family, and the immunoglobulin ϵ region. Saitou and Nei (1986) estimated the gene tree from the combined data of the mtDNA I and mtDNA II fragments, and human and chimpanzee were clustered in their tree. It is clear that we obtain the same clustering for the other two gene trees, although none of them is statistically significant. Recently, however, chimpanzee and gorilla were found to be closer to each other than to human for the case of the involucrin gene (Djlan and Green, 1989). It has been shown that several gene trees may be necessary for obtaining the correct species tree in the case of human, chimpanzee, and gorilla (Saitou and Nei, 1986). Therefore, examination of at least a few more genes seem to be necessary to establish the branching pattern of human, chimpanzee, and gorilla.

Congruence between the mtDNA and nuclear DNA data

It has been shown from the comparison of evolutionary distances for human, chimpanzee, and gorilla that there was a good congruence between the data for mitochondrial DNA and those for nuclear DNA (see Table 2). The range of this congruence can be checked by making a similar comparison for the data of another trio of species (chimpanzee, pygmy chimpanzee, and human).

The proportion (P) of the branching point for chimpanzee and pygmy chimpanzee to that for human and the chimpanzee cluster is given by $P = 2 \times$ D(chimpanzee-pygmy chimpanzee)/[D(human − chimpanzee) + D(human − pygmy chimpanzee)]. As for the nuclear DNA, data from protein electrophoresis (Bruce and Ayala, 1979) and those from DNA-DNA hybridization (Sibley and Ahlquist, 1987; Caccone and Powell, 1989) give P values of 0.30, 0.42, and 0.49, respectively. On the other hand, the P value for the mtDNA II fragment data is 0.35 (see Table 1b), and the restriction-site data of Ferris et al. (1981) gives $P = 0.32$ [this P value is obtained from the evolutionary distances given by Nei et al. (1985)]. Nucleotide sequence data obtained for mitochondrial DNA by Foran et al. (1988) were not used because they sample a noncoding region containing a lot of insertions and deletions are involved. There is a rough agreement between the P values for mtDNA data and those for nuclear DNA.

Hasegawa et al. (1985) analyzed the nucleotide sequences for the mtDNA I fragment, and their value of D(X − human)/D(Z − orangutan) was 0.25 (from their Fig. 2).

Compared with our value of 0.49 (Table 2), this is rather low. To explain this unusual result, Hasegawa et al. (1985, 1987, 1989) suggested the possibility of interspecies transfer (introgression) of mtDNA during the speciation process of human and chimpanzee. However, we did not find a considerable difference between the evolutionary distances involving human, chimpanzee, pygmy chimpanzee, and orangutan for the mitochondrial DNA and those for the nuclear DNA. Therefore, there seems to be no need to postulate the introgression of mtDNA between the human and chimpanzee lineages.

It is possible that the unusual result of Hasegawa et al. (1985) originated from the use of their evolutionary model, which involves the proportion (f) of the variable nucleotide sites. Hasegawa et al. analyzed the sequence data for mtDNA I fragment [Brown et al.'s (1982) data and data for bovine and mouse] and obtained the f values of 0.53 (the average of the f values for the class I and class II sites, weighted with the number of nucleotides for each class), which is slightly larger than the observed f value of 0.52. However, if we use Hayasaka et al.'s (1988) data, which contain nucleotide sequences of seven more species than Hasegawa et al. analyzed, the observed f value becomes 0.63. Clearly, the proportion of variable sites has increased because of the inclusion of more sequence data. Unfortunately, the parameter f in an evolutionary model has been known to give a large effect for the estimation of evolutionary distances (Nei, 1987).

Congruence of molecular data

Without an assumption or calibration of a molecular clock, molecular data alone can give only the "relative" divergence time. This is why we presented the proportions of molecular distances, or D(X – human)/D(Z – orangutan), in Table 2. These values (average 0.44) are close to the corresponding values of 0.45 and 0.47 obtained, respectively, from the DNA-DNA hybridization data of Sibley and Ahlquist (1987:Table 5) and of Caccone and Powell (1989:Table 4). Therefore, there is a good congruence between the nucleotide sequence data and the DNA-DNA hybridization data. Nei (1985) also showed that the estimates of the divergence times of chimpanzee and gorilla from the human lineage obtained from Brown et al.'s (1982) data were close to those of Sibley and Ahlquist (1984).

However, there still exists an inconsistency between these two kinds of molecular data when the branch length estimates of human, chimpanzee, and gorilla are considered. The proportion of D(X – human)/D(Y – gorilla) for the tree of Figure 2 is 0.89 (= 0.00695/0.00783), whereas the corresponding proportions for the DNA-DNA hybridization data are 0.73 and 0.60 for the data of Sibley and Ahlquist (1987) and Caccone and Powell (1989), respectively. Let us decrease D(X – human) by 1.5 times the standard error and increase D(Y – gorilla) by 1.5 times the standard error for the tree of Figure 2. Then the new value of D(X – human)/D(Y – gorilla) becomes 0.72[= (0.00695 – 0.00081)/(0.00783 + 0.00075)], which is now similar to that of Sibley and Ahlquist (1987) but is still considerably greater than that of Caccone and Powell (1989). If we assume a normal distribution for branch length estimates and ignore the correlation between the two estimates, the probability of obtaining such a low value of D(X – human)/D(Y – gorilla) becomes $0.0668^2 = 0.004$. Therefore, the difference of the value of D(X – human)/D(Y – gorilla) for the nucleotide sequence data and that for DNA-DNA hybridization data seem to be statistically significant.

It should also be noted that there is a controversy on the data presented by Sibley and Ahlquist (Marks et al., 1988, 1989; Britten, 1989). Reanalyzing a part of Sibley and Ahlquist's (1987) data, Marks et al. (1988:Table 2) obtained similar evolutionary distances between the pair of human and chimpanzee (1.54) and that of human and gorilla (1.68). This makes the value of D(X – human)/D(Y – gorilla) equal to 0.92, which is now close to that (0.89) for the nucleotide sequence data. Therefore, we may conclude that the problem of the determination of the branching order among human, chimpanzee, and gorilla is yet to be solved at this moment.

ACKNOWLEDGMENTS

I am grateful to Drs. S. Horai and K. Hayasaka for sending me their mitochondrial sequence data in a computer file and to Dr. S. Ueda for discussion. I also thank Dr. M. Cartmill and two anonymous reviewers for their valuable comments. This study was supported by grants-in-aid from the Ministry of Education, Science, and Culture of Japan.

LITERATURE CITED

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

Aquadro CF, Kaplan N, and Risko KJ (1984) An analysis of the dynamics of mammalian mitochondrial DNA sequence evolution. Mol. Biol. Evol. *1:*423–434.

Britten RJ (1989) Comment on a criticism of DNA hybridization measurements. J. Hum. Evol. *18:*163–164.

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

Bruce EJ, and Ayala FJ (1979) Phylogenetic relationships between man and the apes: Electrophoretic evidence. Evolution *33:*1040–1056.

Caccone A, and Powell JR (1989) DNA divergence among hominoids. Evolution *43:*925–942.

Chang LYE, and Slightom JL (1984) Isolation and nucleotide sequence analysis of the β-type globin pseudogene from human, gorilla and chimpanzee. J. Mol. Biol. *180:*767–783.

Collins FS, and Weissman SM (1984) Molecular genetics of human hemoglobin. Prog. Nucleic Acid Res. Mol. Biol. *31:*315–439.

Djlan P, and Green H (1989) Vectorial expansion of the involucrin gene and the relatedness of the hominoids. Proc. Natl. Acad. Sci. USA *86:*8447–8451.

Eck RV, and Dayhoff MO (1966) In MO Dayhoff (ed): Atlas of Protein Sequence and Structure. Silver Springs, MD: National Biomedical Research Foundation.

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

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

Ferris SD, Wilson AC, and Brown WM (1981) Evolutionary tree for apes and humans based on cleavage maps of mitochondrial DNA. Proc. Natl. Acad. Sci. USA *78:*2432–2436.

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

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

Foran DR, Hixson JE, and Brown WM (1988) Comparison of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. Nucleic Acids Res. *16:*5841–5861.

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

Goodman M, Barnabas J, Matsuda G, and Moore GW (1971) Molecular evolution in the descent of man. Nature *233:*604–613.

Harris S, Barrie PA, Weiss ML, and Jeffreys AJ (1984) The primate φβ1 gene. An ancient β-globin pseudogene. J. Mol. Biol. *180:*785–801.

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

Hasegawa M, Kishino H, and Yano T (1987) Man's place in Hominoidea as inferred from molecular clocks of DNA. J. Mol. Evol. *26:*132–147.

Hasegawa M, Kishino H, and Yano T (1989) Estimation of branching dates among primates by molecular clocks of nuclear DNA which slowed down in Hominoidea. J. Hum. Evol. *18:*461–476.

Hayasaka K, Gojobori T, Horai S (1988) Molecular phylogeny and evolution of primate mitochondrial DNA. Mol. Biol. Evol. *5:*626–644.

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

Holmes EC, Pesole G, and Saccone C (1989) Stochastic models of molecular evolution and the estimation of phylogeny and rates of nucleotide substitution in the hominoid primates. J. Hum. Evol. *18:*775–794.

Holmquist R, Miyamoto MM, and Goodman M (1988) Analysis of higher-primate phylogeny from transversion differences in nuclear and mitochondrial DNA by Lake's methods of evolutionary parsimony and operator metrics. Mol. Biol. Evol. *5:*217–236.

Jukes TH, and Cantor CR (1969) Evolution of protein molecules. In HN Munro (ed): Mammalian Protein Metabolism, Vol. 3. New York: Academic Press, pp. 21–132.

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

Kimura M (1983) The Neutral Theory of Molecular Evolution. Cambridge: Cambridge University Press.

Kohne DE, Chiscon JA, and Hoyer BH (1972) Evolution of primate DNA sequences. J. Hum. Evol. *1:*627–644.

Koop BF, Goodman M, Xu P, Chan K, and Slightom JL (1986) Primate η-globin DNA sequences and man's place among the great apes. Nature *319:*234–238.

Lake JA (1987) A rate-independent technique for analysis of nucleic acid sequences: Evolutionary parsimony. Mol. Biol. Evol. *4:*167–191.

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

Li WH (1989) A statistical test of phylogenies estimated from sequence data. Mol. Biol. Evol. *6:*424–435.

Li WH, and Tanimura M (1987) The molecular clock runs more slowly in man than in apes and monkeys. Nature *326:*93–96.

Li WH, Wu CI, and Luo CC (1984) Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. J. Mol. Evol. *21:*58–71.

Maeda N, Bliska JB, and Smithies O (1983) Recombination and balanced chromosome polymorphism suggested by DNA sequence 5′ to the human delta-globin gene. Proc. Natl. Acad. Sci. USA *80:*5012–5016.

Maeda N, Wu CI, Bliska J, and Reneke J (1988) Molecular evolution of intergenic DNA in higher primates: Pattern of DNA changes, molecular clock, and evolution of repetitive sequences. Mol. Biol. Evol. *5:*1–20.

Marks J, Schmid CW, and Sarich VM (1988) DNA hybridization as a guide to phylogeny: Relations of the Hominoidea. J. Hum. Evol. *17:*769–786.

Marks J, Schmid CW, and Sarich VM (1989) Response to Britten. J. Hum. Evol. *18:*165–166.

Miyamoto MM, Slightom JL, and Goodman M (1987) Phylogenetic relations of humans and African apes from DNA sequences in the φη-globin region. Science *238:*369–373.

Nei M (1985) Human evolution at the molecular level. In T Ohta and K Aoki (eds): Population Genetics and Molecular Evolution. Tokyo: Japan Sci. Soc. Press, pp. 41–64.

Nei M (1987) Molecular Evolutionary Genetics. New York: Columbia University Press.

Nei M, Stephens JC, and Saitou N (1985) Methods for computing the standard errors of branching points in an evolutionary tree and their application to molecular data from humans and apes. Mol. Biol. Evol. *2:*66–85.

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

Saitou N (1991) Statistical methods for phylogenetic tree reconstruction. In CR Rao and R Chakraborty (eds.): Handbook of Statistics, Vol. 8: Statistical Methods for Biological and Medical Sciences. New York: North Holland Publishing Co., in press.

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

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

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

Sakoyama Y, Hong KJ, Byun SM, Hisajima H, Ueda S, Yaoita Y, Hayashida H, Miyata T, and Honjo T (1987) Nucleotide sequences of immunoglobulin ϵ genes of chimpanzee and orangutan: DNA molecular clock and hominoid evolution. Proc. Natl. Acad. Sci. USA *84:*1080–1084.

Sarich VM, and Wilson AC (1967) Immunological time scale for hominoid evolution. Science *158:*1200–1204.

Sibley CG, and Ahlquist JE (1987) DNA hybridization evidence of hominoid phylogeny: Results from an expanded data set. J. Mol. Evol. *26:*99–121.

Sokal RR, and Sneath PHA (1963) Principles of Numerical Taxonomy. San Francisco: WH Freeman.

Sourdis J, and Nei M (1988) Relative efficiencies of the maximum parsimony and distance-matrix methods in obtaining the correct phylogenetic tree. Mol. Biol. Evol. *5:*298–311.

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

Takahata N, and Nei M (1985) Gene genealogy and variance of interpopulational nucleotide differences. Genetics *110:*325–344.

Ueda S, Nakai S, Nishida Y, Hisajima H, and Honjo T (1982) Long terminal repeat-like elements flank a human immunoglobulin epsilon pseudogene that lacks introns. EMBO J. *1:*1539–1544.

Ueda S, Watanabe Y, Saitou N, Omoto K, Hayashida H, Miyata T, Hisajima H, and Honjo T (1989) Nucleotide sequences of immunoglobulin epsilon pseudogenes in hominoids and their phylogenetic relationships. J. Mol. Biol. *205:*85–90.

Williams SA, and Goodman M (1989) A statistical test that supports a human/chimpanzee clade based on noncoding DNA sequence data. Mol. Biol. Evol. *6:*325–330.

NOTE ADDED IN PROOF

Recently Hasegawa (1990; *Jpn. J. Genet* 65:243–266) presented an analysis of Hayasaka et al.'s (1988) data, and D(X-human)/D(Z-orangutan) became 0.30 for mtDNA I, which is larger than the value (0.25) obtained by Hasegawa et al. (1985). However, this new estimate is still considerably smaller than our estimate (0.49). Because the method we used in this study for estimating evolutionary distances does not take into account the heterogeneity of evolutionary rate at different nucleotide sites, which does exist in real situation, it is possible that our value of D(Z-orangutan) is slightly underestimated, resulting a relatively large value of D(X-human)/D(Z-orangutan).