Evolutionary Rates of Insertion and Deletion in Noncoding Nucleotide Sequences of Primates

Naruya Saitou and Shintaroh Ueda†*

*Laboratory of Evolutionary Genetics, National Institute of Genetics; and †Department of Anthropology, Graduate School of Science, University of Tokyo

Insertions and deletions are responsible for gaps in aligned nucleotide sequences, but they have been usually ignored when the number of nucleotide substitutions was estimated. We compared six sets of nuclear and mitochondrial noncoding DNA sequences of primates and obtained the estimates of the evolutionary rate of insertion and deletion. The maximum-parsimony principle was applied to locate insertions and deletions on a given phylogenetic tree. Deletions were about twice as frequent as insertions for nuclear DNA, and single-nucleotide insertions and deletions were the most frequent in all events. The rate of insertion and deletion was found to be rather constant among branches of the phylogenetic tree, and the rate (~2.0/kb/Myr) for mitochondrial DNA was found to be much higher than that (~0.2/kb/Myr) for nuclear DNA. The rates of nucleotide substitution were about 10 times higher than the rate of insertion and deletion for both nuclear and mitochondrial DNA.

Introduction

Evolutionary change of nucleotide sequences has been studied mainly from the viewpoint of nucleotide substitutions. Now we have ample knowledge of the pattern and rate of nucleotide substitution for many organisms and genes (e.g., see Nei 1987). Insertions and deletions of nucleotides are, however, not well studied. One reason is that they are considered to occur infrequently in coding regions, because most of them are strongly deleterious. Tajima and Nei (1984) proposed an evolutionary distance that is calculated on the basis of insertions and deletions, and they applied their method to a coding region, but there was a very low increase of the distance over a long period of evolutionary time.

Mutation is the fundamental source for organismal evolution, and there are many types of mutations, such as nucleotide substitution, insertion, deletion, gene duplication, unequal crossing-over, and gene conversion. To understand the tempo and mode of evolution at the nucleotide level, it is important to estimate the spontaneous rate of each mutation type, including insertions and deletions. Graur et al. (1989) analyzed nucleotide sequences of processed pseudogenes of humans and mice and concluded that deletions accumulated faster in rodents than in humans. However, they showed only a relative rate difference between rodents and humans.

Saitou (1992) conducted an analysis of the insertions and deletions for a small data set of nuclear DNA sequences of hominoids and estimated the absolute rate of insertion and deletion to be 0.18/kb/Myr. In the present study, we analyzed insertions and deletions in the nucleotide sequences of primate noncoding regions for nuclear DNA and mitochondrial DNA, by expanding the data set used by Saitou (1992). We show that there is a constancy in the rate of insertion and deletion, and estimate the evolutionary rate of insertion and deletion. Because most of the changes in the noncoding region of DNA are supposed to be neutral (Kimura 1983), our estimates of evolutionary rate can be equated to the spontaneous mutation rate of insertion and deletion.

Material and Methods

We used six sets of nucleotide sequence data, four of them for nuclear DNA and two for mitochondrial DNA. These were as follows (lengths of each region are averages of sequences used): (1) The η-globin ψ is a 6.4-kb fragment containing the η-globin pseudogene in the β-globin gene family. Aligned sequence data for human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*), orangutan (*Pongo pigmaeus*), gibbon (*Hylobates lar*), rhesus macaque (*Macaca mulatta*), and spider monkey (*Ateles geoffroyi*) were taken from figure 3 (positions 1–8740) of Bailey et al. (1992). (2) The β-globin spacer is a 3.1-kb fragment containing a spacer DNA of the β-globin gene family. Aligned se-

Key words: insertions and deletions, molecular clock, evolutionary rate, primates, mutation rate.

Address for correspondence and reprints: Naruya Saitou, Laboratory of Evolutionary Genetics, National Institute of Genetics, 1111 Yata, Mishima 411, Japan.

Mol. Biol. Evol. 11(3):504–512. 1994.

0737-4038/94/1103-0017$02.00

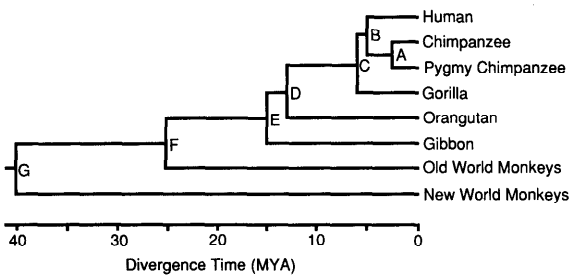

FIG. 1.—Assumed phylogenetic tree of primates. Branching points A–G correspond to 2.5, 5.0, 6.0, 13.0, 15.0, 25.0, and 40.0 Mya, respectively.

A

```
CCGCCCCTTCA: Human
CCGCCCCTTCA: Chimpanzee
CCGCCCCTGCA: Gorilla
CCGCCCC-GCA: Orangutan
CC-CCCCTTCA: Crab-eating macaque
```

B

```
AGGTGGTGGTGGACA: Human
AGGTGGTGGTGGACA: Chimpanzee
AGGTGG---TGGACA: Gorilla
AGGTGG---TGGACA: Orangutan
AGGTGGTGGTAGACA: Crab-eating macaque
```

FIG. 2.—Two examples of the gaps in the aligned sequences (data from Kawamura et al. 1991). *A*, 561–571-bp region of the Ig-α gene. *B*, 607–621-bp region of the Ig-α gene.

quence data for human (R allele), chimpanzee, gorilla, orangutan, rhesus macaque, and spider monkey were taken from Maeda et al. (1988). (3) The Ig-ε3 ψ is a 2.3-kb fragment containing immunoglobulin ε3 processed pseudogene. Aligned sequence data for human, chimpanzee, gorilla, and orangutan were taken from Ueda et al. (1989). (4) The Ig-α noncoding region is a 2.6-kb fragment containing the immunoglobulin-α coding gene. Aligned sequence data for human, chimpanzee, gorilla, orangutan, and crab-eating macaque (*Macaca fascicularis*) were taken from Kawamura et al. (1991). This fragment contains a 1.5-kb noncoding region, and only this region was used for analysis. (5) The mitochondrial DNA D-loop is a 1.1-kb fragment of noncoding mitochondrial DNA containing the so-called D-loop region. Aligned sequence data for human, chimpanzee, pygmy chimpanzee (*Pan paniscus*), gorilla, and orangutan were taken from Foran et al. (1988). (6) The mitochondrial DNA ⅓-genome is a 5.0-kb fragment of mitochondrial DNA containing 11 tRNA genes, 6 protein genes, and some noncoding regions, which roughly corresponds to one-third of the entire mitochondrial genome. Aligned sequence data for human, chimpanzee, pygmy chimpanzee, gorilla, orangutan, and siamang (*Hylobates syndactylus*) were taken from Horai et al. (1992). We used only the noncoding region, which was 0.1 kb in total.

In general, positions and lengths of gaps created in the alignment procedure may vary depending on the gap penalties used. Because the published multiple aligned sequence data used are closely related, however, most of the gaps will not change their positions and lengths drastically when gap penalties are modified. Thus the ambiguity of gap determination is avoided in the present study.

The method used to assign insertions and deletions to a branch of a phylogenetic tree is as follows: We first assumed the phylogenetic tree of higher primates (see fig. 1), which was constructed by considering molecular data from many sources (e.g., see Sibley and Ahlquist 1987; Hayasaka et al. 1988; Ueda et al. 1989; Saitou 1991). Although there are ambiguities in the detailed pattern, in particular the branching times, we believe that the overall picture will not change drastically. On the basis of this tree, the maximum-parsimony principle (see Sneath and Sokal 1973) was applied, to locate insertions and deletions to each branch of the phylogenetic tree. The assumed phylogenetic tree was different for each data set, but all of them are partial trees of figure 1.

The number of nucleotides involved in insertions and deletions was irrelevant in this operation. A contiguous block of nucleotides (a gap) was considered to be created by one event. For example, there were two gaps in the 561–572 bp of Ig-α sequences (fig. 2*A*). In that case, both gaps were unambiguously allocated to the branches for orangutan (a deletion of 1 nucleotide) and

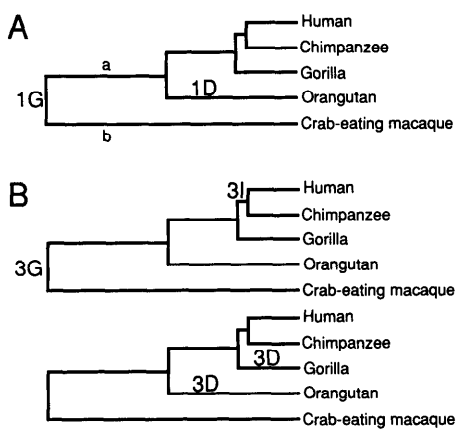

FIG. 3.—Allocation of insertions and deletions to each branch of the phylogenetic tree of fig. 1 by applying the maximum-parsimony principle to the aligned sequence data of fig. 2. "I," "D," and "G" denote insertion, deletion, and gap (insertion or deletion) events, respectively. Numbers before those symbols are the no. of nucleotides involved in each event. *A*, Allocation of events for the data of fig. 2*A*. *B*, Two possible ways of allocation for the data of fig. 2*B*.

Table 1
Length Distribution of Insertions, Deletions, and Gaps

Length (nucleotides)	Insertions	Deletions	Gaps	Total
A. Data set I-A (η-globin pseudogene region):				
1	14 (3.83)	29 (4.33)	19 (0.50)	62 (8.67)
2	5 (1.67)	11	5	21 (1.67)
3	3 (0.33)	4	2	9 (0.33)
4	3 (0.33)	7 (0.33)	3	13 (0.67)
5	1 (0.33)	3 (0.33)	2	6 (0.67)
6	2	3	2	7
7	. . .	1	. . .	1
8	1	1	. . .	2
11	2	. . .	1	3
12	1	. . .	1	2
13	2	1	. . .	3
14	1	1	. . .	2
18	. . .	1	. . .	1
23	. . .	1	. . .	1
24	. . .	3	. . .	3
26	. . .	1	. . .	1
28	. . .	1	. . .	1
38	1	. . .	. . .	1
76	(0.50)	(0.50)	. . .	(1.00)
331	1	. . .	. . .	1[a]
338	. . .	. . .	1	1[a]
339	. . .	. . .	1	1[a]
558	. . .	. . .	1	1[b]
Total	37 (7.00)	68 (5.50)	38 (0.50)	143 (13.00)
B. Data set I-B (β-globin spacer region):				
1	3 (0.50)	5 (1.50)	12 (0.50)	20 (2.50)
2	1 (0.50)	5 (1.00)	4 (0.50)	10 (2.00)
3	. . .	3 (1.00)	1	4 (1.00)
4	. . .	2	1	3
5	. . .	(1.50)	1	1 (1.50)
6	. . .	1 (0.25)	. . .	1 (0.25)
7	. . .	(1.00)	. . .	(1.00)
8	1	(0.25)	. . .	1 (0.25)
9	. . .	1 (0.50)	1	2 (0.50)
12	. . .	. . .	1	1
14	. . .	. . .	1	1
15	. . .	. . .	1	1
20	. . .	. . .	1	1
45	. . .	. . .	1	1
Total	5 (1.00)	17 (7.00)	25 (1.00)	47 (9.00)
C. Data set I-C (immunoglobulin-α gene noncoding region):				
1	. . .	6	8	14
2	. . .	. . .	2	2
3	(0.5)	(1.0)	(0.5)	(2.0)
9	1	. . .	1	2
10	. . .	. . .	1	1
11	. . .	1	. . .	1
12	. . .	. . .	1	1
13	. . .	1	. . .	1
15	1	. . .	. . .	1
30	. . .	1	. . .	1
42	. . .	1	. . .	1
Total	2 (0.5)	10 (1.0)	13 (0.5)	25 (2.0)

Table 1 (*Continued*)

Length (nucleotides)	Insertions	Deletions	Gaps	Total
D. Data set I-D (immunoglobulin ε pseudogene region):				
1	1 (0.40)	3 (0.40)	2 (0.40)	6 (1.20)
2	(0.33)	1 (0.73)	1	2 (1.07)
3	. . .	. . .	1	1
4	1	(0.33)	1	2 (0.33)
5	(0.33)	. . .	. . .	(0.33)
6	(0.20)	(0.33)	. . .	(0.53)
7	. . .	(0.60)	(0.20)	(0.80)
8	(0.20)	(0.20)	. . .	(0.40)
9	(0.33)	. . .	(0.33)	(0.67)
11	(0.33)	. . .	(0.33)	(0.67)
16	1	. . .	. . .	1
Total	3 (2.13)	4 (2.60)	5 (1.27)	12 (6.00)
E. Data set II-A (mitochondrial DNA D-loop region):				
1	5 (0.75)	8 (1.00)	9 (0.75)	22 (2.50)
2	(0.25)	2 (0.25)	4	6 (0.50)
3	. . .	. . .	1	1
4	. . .	1	2	3
16	. . .	. . .	1	1
27	. . .	. . .	1	1[c]
38	. . .	. . .	1	1
71	. . .	. . .	1	1[c]
Total	5 (1.00)	11 (1.25)	20 (0.75)	36 (3.00)
F. Data set II-B (mitochondrial DNA ⅓-genome region):				
1	2 (0.33)	2 (0.67)	3 (0.33)	7 (1.33)
2	. . .	(0.33)	(0.33)	(0.67)
4	1	. . .	. . .	1
7	1	. . .	. . .	1
16	1	. . .	. . .	1
Total	5 (0.33)	2 (1.00)	3 (0.67)	10 (2.0)

NOTE.—Values in parentheses are those with multiple possibilities.
[a] *Alu* element.
[b] L1 element.
[c] Region where the alignment was impossible to establish.

for crab-eating macaque (fig. 3*A*). It should be noted that the root of the tree of figure 3*A* was ignored in that operation. It could not be determined whether a gap at the crab-eating macaque sequence in figure 2*A* was an insertion (at branch "a" of fig. 3*A*) or a deletion (at branch "b" of fig. 3*A*), and it was designated "1G" (1-nucleotide gap) in figure 3*A*. In this paper, "gap" means an event for which it cannot be determined whether an insertion or a deletion occurred.

When there was more than one possibility for allocation of gaps to branches, all possibilities were considered, and gaps were allocated equally to corresponding branches. For example, there were multiple gaps in the 607–621-bp region of Ig-α sequences (fig. 2*B*). In that case, there were two possible patterns of insertion and deletion (fig. 3*B*). One possibility was that the lineage to the common ancestor of human and chimpanzee experienced an insertion of 3 nucleotides, and a 3-nucleotide-long gap occurred in the crab-eating macaque lineage or in the lineage to the common ancestor of human, chimpanzee, gorilla, and orangutan (*upper tree* of fig. 3*B*). Another possibility was that both gorilla and orangutan lineages experienced a deletion of 3 nucleotides (*lower tree* of fig. 3*B*). We allocated all these possible insertions and deletions to the corresponding branches with a probability of one-half. This allocation procedure is simpler than that of Saitou (1992).

Results and Discussion

Distribution of the Lengths of Insertions and Deletions

Table 1 shows the distribution of the lengths of insertions, deletions, and gaps for each data set. Values listed are the numbers of insertions, deletions, and gaps

Table 2
Summary of the Number of Insertions, Deletions, and Gaps

Data Set	Insertions	Deletions	Gaps	Total	Mean Length
Nuclear DNA:					
η-Globin ψ[a]	36 (43.0)	68 (73.5)	35 (35.5)	139 (152)	4.3 (4.6)
β-Globin spacer	5 (6.0)	17 (24.0)	25 (26.0)	47 (56)	4.4 (4.2)
Ig α	2 (2.5)	10 (11.0)	13 (13.5)	25 (27)	6.8 (6.5)
Ig ε ψ	3 (5.1)	4 (6.6)	5 (6.3)	12 (18)	3.1 (3.8)
Total	46 (56.6)	99 (115.1)	78 (81.3)	223 (253)	4.5 (4.6)
Mitochondrial DNA:					
D-loop	5 (6.0)	11 (12.3)	20 (20.7)	36 (39)	5.6 (5.2)
⅓-Genome	5 (5.3)	2 (3.0)	3 (3.7)	10 (12)	3.4 (3.1)
Total	10 (11.3)	13 (15.3)	23 (24.4)	46 (51)	5.1 (4.7)

NOTE.—Values are those for unambiguously determined events, and those in parentheses include ambiguously determined events.
[a] Alu and L1 repeats were excluded.

for each branch, that were determined unambiguously, and values in parentheses are those with multiple possibilities. A total of 308 gap-making events were extracted from the six data sets, of which 273 gaps were unambiguously determined.

Four long gaps were found in η-globin pseudogene (those marked "a" and "b" in table 1*A*), and they are either Alu or L1 elements (Bailey et al. 1992). Because those gaps were probably created by a mechanism different from that responsible for shorter gaps, we eliminated them from the following analysis.

There were two regions in the gorilla sequence of mitochondrial DNA D-loop where the alignment was very difficult to determine (Foran et al. 1988). We treated these regions as blocks, which resulted in two gaps with lengths of 27 and 71 nucleotides (those marked "c" in table 1*E*). Probably more than one gap-creating event occurred in each block. We do not know how many more occurred, yet those data were included in the following analysis.

Single-nucleotide insertions and deletions were the most frequent in all the data sets, and the frequency of gaps decreases as the length increases. A summary of the number of insertions and deletions is presented in table 2. Deletions are about twice as frequent as insertions, for nuclear DNA. Deletions occur at a slightly higher frequency than insertions, in mitochondrial DNA; however, the difference is not as large as that for nuclear DNA.

de Jong and Rydén (1981) compared human hemoglobin variants with hemoglobin pseudogene sequences and showed that deletions outnumbered insertions. Graur et al. (1989) also found the same tendency by analyzing human and rodent processed pseudogene sequences. The present results seem to confirm previous studies, though the ratio of deletions to insertions varies. de Jong and Rydén (1981) proposed a model that explained the excess of spontaneous deletions to insertions.

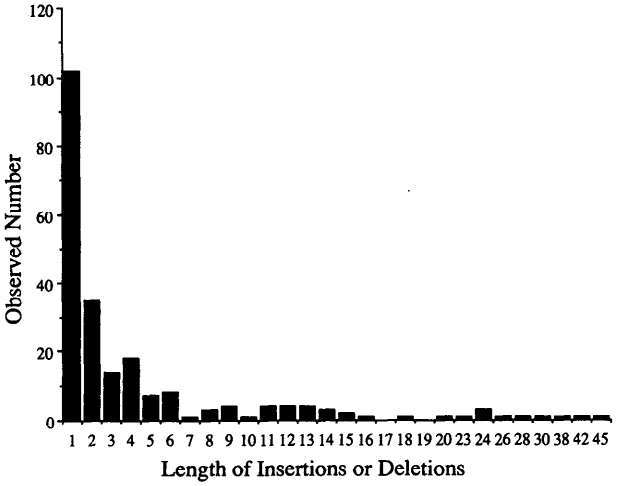

FIG. 4.—Distribution of the lengths of insertions or deletions for nuclear DNA. Results for all four data sets are combined.

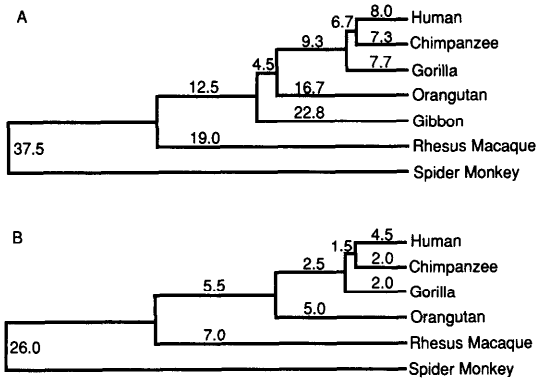

FIG. 5.—Insertions/deletions of β-globin gene region mapped to the given phylogenetic tree. *A*, η-Globin pseudogene region (data from Bailey et al. 1992). *B*, β-Globin gene spacer region (data from Maeda et al. 1988).

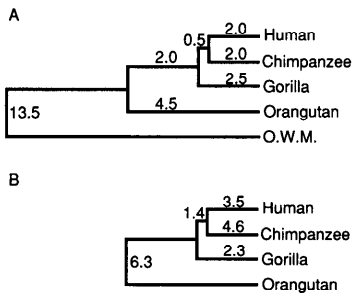

FIG. 6.—Insertions/deletions of immunoglobulin gene region mapped to the given phylogenetic tree. *A*, Ig-α gene noncoding region (data from Kawamura et al. 1991). *B*, Ig-ε pseudogene region (data from Ueda et al. 1989).

The mean lengths of insertions or deletions were computed for each data set and are shown at the rightmost column of table 2. It is interesting that the mean is rather uniform over all regions, and the overall averages for nuclear DNA and mitochondrial DNA are 4.5 bp and 5.1 bp, respectively. Values in parentheses in table 2 include ambiguously determined insertions and deletions. The mean lengths are not much different from those for unambiguously determined insertions and deletions, and the overall averages for nuclear DNA and mitochondrial DNA were 4.6 bp and 4.7 bp, respectively.

Figure 4 shows the distribution of the lengths of insertions and deletions for nuclear DNA (results of all four data sets were combined). Only the unambiguously determined insertions, deletions, or gaps were used. About 50% (102 of 223) of the total events were single-nucleotide gaps (insertions or deletions), and the frequency of gap events decreases as the length increases. A high occurrence of single-nucleotide gaps was also observed by Graur et al. (1989).

Golenberg et al. (1993) analyzed about 0.9 kb of chloroplast noncoding nucleotide sequence for nine monocot plants and found 23 single-nucleotide insertions or deletions from 56 events. Mo et al. (1991) analyzed the mutation pattern of the *rpsL* gene of *Escherichia coli*. They found 12 single-nucleotide insertions or deletions, while gaps longer than 1 nucleotide occurred 13 times (47 insertion sequences were excluded in that value). Therefore, a high prevalence of single-nucleotide gaps seems to be a common phenomenon both in eukaryote and prokaryote nucleotide evolution.

Rates of Insertion and Deletion for Nuclear and Mitochondrial DNAs

Insertions and/or deletions were mapped to the phylogenetic tree of primates (fig. 1) for each data set, and the results are shown in figures 5–7. Figure 5 shows those for η-globin ψ and β-globin spacer regions, figure 6 for Ig-ε3 ψ and Ig-α noncoding regions, and figure 7 for mitochondrial DNA D-loop and ⅓-genome regions. Insertions and deletions that were both ambiguously and unambiguously determined were included in this analysis.

The number of gaps is approximately proportional to the length (evolutionary time) of each branch of the trees. This suggests that the evolutionary rate of insertion and deletion is more-or-less constant over primate evolution. We therefore estimated the rates of insertion and deletion as follows.

The evolutionary rate of insertion and deletion was estimated by using two methods. The first one is that of Saitou (1992), in which the total number of insertions and deletions is divided by the total branch length (in terms of evolutionary time) of the assumed phylogenetic tree and by the length of the nucleotide sequences compared. Table 3 shows the rates thus obtained. It is interesting that rates among four nuclear DNA data sets are more-or-less constant, with a range of 0.14–0.24/kb/Myr. If we take the average for those values, weighted by the length of nucleotides compared, it becomes 0.17/kb/Myr.

The same is true for mitochondrial DNA. The rates for D-loop region and for ⅓-genome region were estimated to be 1.81 and 2.12/kb/Myr, respectively, and the weighted mean became 1.84/kb/Myr (see table 3). Thus the rate of insertion and deletion in mitochondrial DNA seems to be >10 times higher than that in nuclear DNA.

Saitou (1992) analyzed four sets of noncoding nucleotide sequences of hominoid nuclear DNA and estimated the evolutionary rate of insertion and deletion to be 0.18/kb/Myr. That estimate is close to the corresponding rate (0.17/kb/Myr) obtained by the present study, though the sequence data that were used overlapped with those used for the present study.

A simple regression analysis was used for the second method. In this case, each branch of a tree was plotted

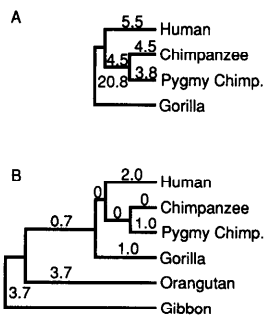

FIG. 7.—Insertions/deletions of mitochondrial DNA mapped to the given phylogenetic tree. *A*, D-loop region (data from Foran et al. 1988). *B*, ⅓-genome region (data from Horai et al. 1992).

separately. The abscissa is evolutionary time (with the unit of Myr), and the ordinate is the number of insertions and deletions per kilobase of a branch (see fig. 8). Because each branch of a tree evolved independently from each other, the observations are independent. This justifies the regression analysis. Because there should be neither insertions nor deletions when the evolutionary time is zero, the regression through the origin (Zar 1984) was calculated. In this case, the regression coefficient or the evolutionary rate (b) of insertion and deletion is estimated by

$$b = \Sigma x_i y_i / \Sigma x_i^2, \qquad (1)$$

where x_i and y_i are branch length and number of insertions and deletions per kilobase for branch i, respectively.

Evolutionary rates thus obtained for nuclear and mitochondrial DNA are presented, with their standard errors, in the right-most column of table 3. Standard errors for those estimates were unexpectedly low, which suggests constancy for the evolutionary rate of insertion and deletion. The rate for the Ig-α gene was considerably higher than those for the remaining three nuclear DNA data, and the difference was statistically significant. This seems to suggest that the rate of insertion and deletion varies from gene to gene in the nuclear DNA. Because the data set used in the present study was not extensive and because the difference of rates between nuclear and mitochondrial DNA data was much larger than that among the nuclear DNA data, however, we combined the four data sets for nuclear DNA, as well as two for mitochondrial DNA, to elucidate a general picture (see fig. 8). Evolutionary rates for nuclear and mitochondrial DNA thus became 0.15 ± 0.01 and 2.27 ± 0.21/kb/Myr, respectively. It is clear that the evolutionary rate for mitochondrial DNA is much higher than that for nuclear DNA. The estimates obtained from regression analysis are not much different than those obtained by the first method.

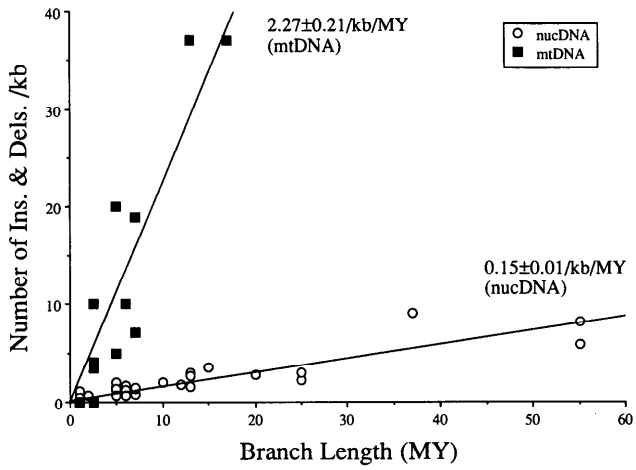

FIG. 8.—Comparison of the evolutionary rates of insertions and deletions between nuclear and mitochondrial DNA.

In this analysis, the maximum-parsimony principle was used both for multiple alignments and for allocation of gaps. This may cause underestimation of the evolutionary rate, when we compare divergent sequences, where multiple insertions and deletions may occur at the same positions, as in the case of multiple nucleotide substitutions. Unlike the various correction methods for estimating the number of nucleotide substitutions, however, there is, so far, no method for correcting for multiple insertions and deletions.

We thus examined the effect of branch lengths on the estimation of the rates of insertion and deletion for nuclear DNA. Table 4 shows the rates estimated for var-

Table 3
Estimates for the Evolutionary Rates of Insertion and Deletion

Data Set	No. of Gap Events	Length (kb)	Total Time (Myr)	Rate (/kb/Myr)	
				Method 1	Method 2[a]
Nuclear DNA:					
η-Globin ψ	152	6.4	144	0.16	0.12 ± 0.04 (10)
β-Globin spacer	56	3.2	129	0.14	0.14 ± 0.01 (8)
Ig α	27	1.5	74	0.24	0.24 ± 0.01 (6)
Ig ε ψ	18	2.3	37	0.21	0.16 ± 0.03 (4)
Overall	253	13.4		0.17	0.15 ± 0.01 (31)
Mitochondrial DNA:					
D-loop	39	1.1	19.5	1.81	2.01 ± 0.38 (4)
⅓-Genome	12	0.1	56.5	2.12	2.31 ± 0.26 (8)
Overall	51	1.2		1.84	2.27 ± 0.21 (13)

[a] Regression coefficient through the origin ± its standard error. Values in parentheses are degrees of freedom.

Table 4
Comparison of the Rates of Insertion and Deletion for Nuclear DNA at Each Branch of the Phylogenetic Tree

Branch (Length)[a]	Rate for (/kb/Myr)				
	η-Globin	β-Globin	Ig-α	Ig-ε3	Average
C-B (1)	1.05	0.47	0.33	0.61	0.75
B-human (5)	0.25	0.28	0.27	0.30	0.27
B-A-chimpanzee (5)	0.23	0.13	0.27	0.40	0.24
C-gorilla (6)	0.20	0.10	0.28	0.17	0.18
D-C (7)	0.21	0.11	0.19	. . .	0.18
D-orangutan (13)	0.20	0.12	0.23	. . .	0.18
D-E-F-Old World monkey (37)	0.13	0.11	0.24	. . .	0.14
F-G-New World monkey (55)	0.11	0.15	. . .	. . .	0.12

[a] Branches of fig. 1. Length is in Myr.

ious branches of the phylogenetic tree of figure 1. There is indeed a tendency for evolutionary rates of longer branches to be slower than those for shorter ones. Therefore, the true rate of insertion and deletion might be somewhat greater than the rate (0.15/kb/Myr) obtained by considering all branches. A similar observation was obtained by Golenberg et al. (1993), who compared the rates within and among tribes of plants. More sequence data will be necessary to clarify this point.

Saitou (1992) showed that the evolutionary rate of insertion and deletion for the branch C-D (see fig. 1) was much higher than those for the remaining branches of the phylogenetic tree involving human, chimpanzee, gorilla, and orangutan. The same tendency was also observed in the present study; however, the high estimate of the rate for the branch C-D depends on the branch-length estimation. If the real branch length is 2 Myr instead of 1 Myr, as assumed in the present study, the considerably high rate along branch C-D decreases.

Implications of the Evolutionary Rates of Insertion and Deletion

Li and Sadler (1991) compared allelic differences of human nuclear DNA sequences and found 4.4 gaps in 23.4 kb of 3′ and 5′ untranslated regions of exons. If we use our estimate (~0.15–0.17/kb/Myr) for the evolutionary rate of gaps for nuclear DNA, divergence time of two human alleles becomes $(4.4/23.4)/(2 \times \sim 0.15\text{–}0.17) = \sim 0.55\text{–}0.63$ Myr, under the assumption of constancy of the evolutionary rate. This divergence time is consistent with the estimate for the coalescence time (~0.8 Myr) of neutral nuclear genes of humans (Takahata 1993). This supports our result that the evolutionary rate of insertion and deletion is more-or-less constant.

Saitou (1991) estimated the number of nucleotide substitutions between human and chimpanzee to be 0.0139/nucleotide site for noncoding nuclear DNA. If we assume the divergence time between these two species to be 5 million years, then the evolutionary rate of nucleotide substitution becomes 1.39×10^{-9}/nucleotide site/year, or 1.39/kb/Myr. Therefore, nucleotide substitutions occur about eight to nine times more frequently than insertions and deletions in noncoding regions of primate nuclear DNA.

The rate of insertion and deletion for mitochondrial DNA was estimated to be ~1.8–2.3/kb/Myr. Thus the rate of insertion and deletion for mitochondrial DNA is much higher (~11–15 times) than that for nuclear DNA. Kondo et al. (1993) estimated the rate of synonymous substitution to be 23.7/kb/Myr, which is ~17 times higher than that for nuclear DNA. Therefore, nucleotide substitutions again occur more frequently (~10–13 times) than insertions and deletions in noncoding regions of primate mitochondrial DNA.

These clear-cut rate differences between substitution (high) and insertion and/or deletion (low) and between mitochondrial DNA (high) and nuclear DNA (low) suggest a common mechanism for higher evolutionary rate, for mitochondrial DNA than for nuclear DNA, of nucleotide substitution and of insertion and deletion. It is possible that the low fidelity of DNA replication in mammalian mitochondrial DNA affects the rate of both substitution and insertion and/or deletion.

If this conjecture is true for primate or mammalian genomes in general, then we expect to observe a similar evolutionary rate of insertion and deletion for nuclear and mitochondrial DNA of *Drosophila,* in which the rate of nucleotide substitution is known to be more-or-less the same between nuclear and mitochondrial DNA (Powell et al. 1986; Sharp and Li 1989). Our prediction awaits the arrival of abundant nucleotide sequence data for *Drosophila.* In conclusion, as far as primate DNA is

concerned, the rate of nucleotide substitution seems to be ~10 times higher than that of insertion and deletion, and both rates are >10 times higher in mitochondrial DNA than in nuclear DNA.

Acknowledgments

The authors thank an anonymous reviewer for his comments and Mrs. T. Kawamoto for technical assistance. This study was supported by grants-in-aid from the Ministry of Education, Science, and Culture, Japan.

LITERATURE CITED

BAILEY, W., K. HAYASAKA, C. G. SKINNER, S. KEHOE, L. C. SIEU, J. L. SLIGHTON, and M. GOODMAN. 1992. Reexamination of the African hominoid trichotomy with additional sequences from the primate β-globin cluster. Mol. Phylogenet. Evol. **1**:97–135.

DE JONG, W. W., and L. RYDÉN. 1981. Causes of more frequent deletions than insertions in mutations and protein evolution. Nature **290**:157–159.

FORAN, D. R., J. E. HIXSON, and W. M. BROWN. 1988. Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. Nucleic Acids Res. **16**:5841–5861.

GOLENBERG, E. M., M. T. CLEGG, M. L. DURBIN, J. DOEBLEY, and D. P. MA. 1993. Evolution of a noncoding region of the chloroplast genome. Mol. Phylogenet. Evol. **2**:52–64.

GRAUR, D., Y. SHUALI, and W.-H. LI. 1989. Deletions in processed pseudogenes accumulate faster in rodents than in humans. J. Mol. Evol. **28**:279–285.

HAYASAKA, K., T. GOJOBORI, and S. HORAI. 1988. Molecular phylogeny and evolution of primate mitochondrial DNA. Mol. Biol. Evol. **5**:626–644.

HORAI, S., Y. SATTA, K. HAYASAKA, R. KONDO, T. INOUE, T. ISHIDA, S. HAYASHI, and N. TAKAHATA. 1992. Man's place in hominoidea revealed by mitochondrial DNA genealogy. J. Mol. Evol. **35**:32–43.

KAWAMURA, S., H. TANABE, Y. WATANABE, K. KUROSAKI, N. SAITOU, and S. UEDA. 1991. Evolutionary rate of immunoglobulin alpha noncoding region is greater in hominoids than in Old World monkeys. Mol. Biol. Evol. **8**:743–752.

KIMURA, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

KONDO, R., S. HORAI, Y. SATTA, and N. TAKAHATA. 1993. Evolution of hominoid mitochondrial DNA with special reference to the silent substitution rate over the genome. J. Mol. Evol. **36**:517–531.

KOOP, B. F., M. GOODMAN, P. XU, K. CHAN, and J. L. SLIGHTON. 1986. Primate η-globin DNA sequences and man's place among the great apes. Nature **319**:234–238.

LI, W.-H., and L. A. SADLER. 1991. Low nucleotide diversity in man. Genetics **129**:513–523.

MAEDA, N., C.-I. WU, J. BLISKA, and J. RENEKE. 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.

MIYAMOTO, M. M., J. L. SLIGHTON, and M. GOODMAN. 1987. Phylogenetic relations of humans and African apes from DNA sequences in the ψη-globin region. Science **238**:369–373.

MO J.-Y., H. MAKI, and M. SEKIGUCHI. 1991. Mutational specificity of the *dnaE173* mutator associated with a defect in the catalytic subunit of DNA polymerase III of *Escherichia coli*. J. Mol. Biol. **222**:925–936.

NEI, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

POWELL, J. R., A. CACCONE, G. D. AMATO, and C. YOON. 1986. Rates of nucleotide substitution in *Drosophila* mitochondrial DNA and nuclear DNA are similar. Proc. Natl. Acad. Sci. USA **83**:9090–9093.

SAITOU, N. 1991. Reconstruction of molecular phylogeny of extant hominoids from DNA sequence data. Am. J. Phys. Anthropol. **84**:75–85.

———. 1992. Evolutionary rate of insertions and deletions in nucleotide sequences. Pp. 335–343 *in* N. TAKAHATA, ed. Population paleo-genetics: proceedings of the Seventeenth Taniguchi International Symposium on Biophysics. Japan Scientific Societies Press, Tokyo.

SHARP, P. M., and W.-H. LI. 1989. On the rate of DNA sequence evolution in *Drosophila*. J. Mol. Evol. **28**:398–402.

SIBLEY, C. G., and J. E. AHLQUIST. 1987. DNA hybridization evidence of hominoid phylogeny: results from an expanded data set. J. Mol. Evol. **26**:99–121.

SNEATH, P. H. P., and R. SOKAL. 1973. Numerical taxonomy. W. H. Freeman, San Francisco.

TAJIMA, F., and M. NEI. 1984. Estimation of evolutionary distance between nucleotide sequences. Mol. Biol. Evol. **1**: 269–285.

TAKAHATA, N. 1993. Allelic genealogy and human evolution. Mol. Biol. Evol. **10**:2–22.

UEDA, S., Y. WATANABE, N. SAITOU, K. OMOTO, H. HAYASHIDA, T. MIYATA, H. HISAJIMA, and T. HONJO. 1989. Nucleotide sequences of immunoglobulin epsilon pseudogenes in hominoids and their phylogenetic relationships. J. Mol. Biol. **205**:85–90.

ZAR, J. H. 1984. Biostatistical analysis, 2d ed. Prentice-Hall, Englewood Cliffs, N.J.

SHOZO YOKOYAMA, reviewing editor

Received November 18, 1993

Accepted January 21, 1994