Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 1080–1084, February 1987
Genetics

Nucleotide sequences of immunoglobulin ε genes of chimpanzee and orangutan: DNA molecular clock and hominoid evolution

YASUHIKO SAKOYAMA*, KYONG-JA HONG*†, SI MYUNG BYUN*‡, HIROSHI HISAJIMA*§, SHINTAROH UEDA¶, YOSHIO YAOITA*‖, HIDENORI HAYASHIDA**, TAKASHI MIYATA**, AND TASUKU HONJO*‖

*Department of Genetics, Osaka University School of Medicine, Osaka 560, Japan; ¶Department of Anthropology, Faculty of Science, The University of Tokyo, Tokyo 113, Japan; and **Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, Japan

Communicated by Motoo Kimura, October 30, 1986

ABSTRACT **To determine the phylogenetic relationships among hominoids and the dates of their divergence, the complete nucleotide sequences of the constant region of the immunoglobulin ε-chain ($C_{\varepsilon 1}$) genes from chimpanzee and orangutan have been determined. These sequences were compared with the human ε-chain constant-region sequence. A molecular clock (silent molecular clock), measured by the degree of sequence divergence at the synonymous (silent) positions of protein-encoding regions, was introduced for the present study. From the comparison of nucleotide sequences of α_1-antitrypsin and β- and δ-globin genes between humans and Old World monkeys, the silent molecular clock was calibrated: the mean evolutionary rate of silent substitution was determined to be 1.56×10^{-9} substitutions per site per year. Using the silent molecular clock, the mean divergence dates of chimpanzee and orangutan from the human lineage were estimated as 6.4 ± 2.6 million years and 17.3 ± 4.5 million years, respectively. It was also shown that the evolutionary rate of primate genes is considerably slower than those of other mammalian genes.**

It has been well documented that amino acid sequences in protein molecules undergo evolutionary changes that are nearly constant with respect to geological time (1–5). Recognition of such clock-like behavior for immunological distances measured by the microcomplement fixation method led Sarich and Wilson (6, 7) to suggest phylogenetic relationships among primates that differed drastically from those that had been accepted by anthropologists in both the tree topology and timing of the divergence of man and apes. The new phylogeny has been supported by accumulating evidence based on comparison of amino acid sequences of different species, on measurement of nucleotide differences by DNA annealing techniques (see refs. 8 and 9 for review), and on comparison of restriction endonuclease cleavage maps of mitochondrial DNAs (10). We attempted to determine the order of branching nodes by analyses of a young pseudogene of the immunoglobulin ε-chain constant region (C_ε) (11).

Miyata *et al.* (12) showed that the synonymous (silent) positions of protein-encoding regions evolved at high rates that were similar for different genes. It was also shown that the silent substitutions occurred with a regular pace during evolution (13, 14). These results suggest that a universal molecular clock (silent molecular clock) with a rapid pace is applicable for many nuclear genes. Thus, comparison of nucleotide sequences would provide direct information as to the date of divergence among primates. The phylogenetic tree among primates has been constructed by direct comparison of nucleotide sequences of the mitochondrial DNA (15, 16). There is, however, no firm data supporting the clock-like behavior of mitochondrial genes.

We report here the nucleotide sequences of the $C_{\varepsilon 1}$ genes from chimpanzee and orangutan and compare them with the human $C_{\varepsilon 1}$ genes, and estimate the time of hominoid evolution using this silent molecular clock.

MATERIALS AND METHODS

Materials. Restriction endonucleases (*Bam*HI, *Sma* I, *Bgl* II, *Sac* I, *Acc* I, *Pvu* II, *Hin*fI, *Ava* II, *Alu* I, *Hae* III, *Taq* I, *Sau*3A, and *Hap* II) were obtained from Takara Shuzo (Kyoto, Japan). *Bst*EII, *Rsa* I, and *Hga* I were purchased from New England Biolabs. T4 DNA ligase, DNA polymerase (Klenow fragment), bacterial alkaline phosphatase, and M13 sequencing kit were from Takara Shuzo. [α-^{32}P]dCTP (≈3000 Ci/mmol; 1 Ci = 37 GBq) was from Amersham and [γ-^{32}P]ATP (≈7000 Ci/mmol) was from New England Nuclear.

Cloning and Characterization of the $C_{\varepsilon 1}$ Genes. High molecular weight DNAs of chimpanzee and orangutan were prepared from blood samples as described (11). Sequences from the $C_{\varepsilon 1}$ genes of chimpanzee and orangutan were identified in 2.7-kilobase (kb) *Bam*HI fragments (11), which were fractionated by electrophoresis in a 0.7% agarose gel. To clone the small *Bam*HI fragment (2.7 kb), we have constructed a derivative of Charon 28 vector. The plasmid $pS_{\gamma 2a}$, which contained the 5.5-kb *Bam*HI fragment of the mouse $S_{\gamma 2a}$ region (17) in pBR322, was digested with *Bgl* II. Since the $pS_{\gamma 2a}$ plasmid has a single *Bgl* II site in the $S_{\gamma 2a}$ region, the resultant 9.6-kb *Bgl* II fragment contains the 4.3-kb pBR322 sequence sandwiched by the $S_{\gamma 2a}$ sequences. The *Bgl* II fragment was cloned into Charon 28. The derivative phage designated Charon 28HS was cleaved with *Bam*HI to remove the pBR322 sequence. The phage arms were elongated 5.3 kb by the addition of the $S_{\gamma 2a}$ sequence. DNA fragments were ligated with the Charon 28HS vector with T4 DNA ligase, and these recombinant phage DNAs were packaged into coat proteins *in vitro*. Phage plaques were screened using the human $C_{\varepsilon 1}$ gene as probe. DNA sequences were determined by the methods of Maxam and Gilbert (18), and the dideoxynucleotide chain-termination method using M13-mp10 and -mp11 (19).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: C_ε, constant region of the immunoglobulin ε chain.

†Present address: Catholic Medical College, Seoul 135, Korea.
‡Present address: The Korea Advanced Institute of Science and Technology, Seoul, Korea.
§Present address: Fujisawa Pharmaceutical Co., Ltd., Tsukuba Exploratory Research Laboratories, Ibaragi 300-26, Japan.
‖Present address: Department of Medical Chemistry, Kyoto University Faculty of Medicine, Sakyo-ku, Kyoto 606, Japan.

RESULTS AND DISCUSSION

Cloning and Nucleotide Sequence Determination of the $C_{\varepsilon 1}$ Genes of Chimpanzee and Orangutan. Genomic DNAs of chimpanzee and orangutan contained two *Bam*HI fragments that hybridized with the human $C_{\varepsilon 1}$ gene: 2.7 kb and 8.0 kb in chimpanzee; 2.7 kb and 7.0 kb in orangutan. The 2.7-kb bands are the active $C_{\varepsilon 1}$ gene fragments (11). The 8.0- and 7.0-kb bands were shown to correspond to the $C_{\varepsilon 3}$ gene, the processed pseudogene (11). The fractionated 2.7-kb *Bam*HI fragments containing the $C_{\varepsilon 1}$ gene were cloned from chimpanzee and orangutan DNAs. The isolated clones from chimpanzee and orangutan were named Ch·PTR·IgC$_{\varepsilon 1}$ and Ch·PPY·IgC$_{\varepsilon 1}$, respectively.

The restriction maps and the sequencing strategies of the 2.7-kb *Bam*HI fragments containing the $C_{\varepsilon 1}$ genes of chimpanzee and orangutan are shown in Fig. 1. The basic structures of these $C_{\varepsilon 1}$ genes agreed with that of the human $C_{\varepsilon 1}$ gene: the coding region of the $C_{\varepsilon 1}$ gene from each species was divided into four exons (C_{H1}, C_{H2}, C_{H3}, and C_{H4}) by three introns at the same positions as the human gene. The coding sequences ended with the termination codon TGA, and the polyadenylylation signal (AATAAA) occurred at the similar position.

Comparison of the $C_{\varepsilon 1}$ Gene Sequences of Chimpanzee and Orangutan with That of the Human. The nucleotide sequences thus determined were compared with that of the human $C_{\varepsilon 1}$ gene reported by us and others (20, 21). Fig. 2 shows the alignment of the sequences of the three species. The discrepancies between the two human sequences were found at nine sites as shown by asterisks (thymidine to cytidine at positions 35, 44, 1571, and 1998; adenosine to thymidine at position 607; thymidine to guanosine at positions 1549 and 1643; AC to GT at positions 1896 and 1897; guanosine deletion at position 518). These sequence discrepancies could be explained by genetic polymorphism or by sequencing errors. We adopted our sequence data because our sequence has fewer mutations than the sequence of either orangutan or chimpanzee except for position 1998.

The degree of nucleotide difference (K_S) at the silent positions of protein encoding regions and that (K_A) at the amino acid-changing positions have been calculated by the method described in ref. 22. In calculating the nucleotide difference (K_N) of noncoding regions, gaps have been counted as substitutions regardless of their lengths. The nucleotide difference K has been corrected for superimposed substitutions by $K^C = -(3/4)\ln[1 - (4/3)K]$ (5, 23). The calculated values of K_S^C, K_A^C, and K_N^C for each genetic segment are shown in Table 1. As previously indicated by comparison between the $C_{\varepsilon 1}$ genes of mouse and man (24), the K_A^C values of the C_{H4} exons are generally the lowest, although the sequences of human and chimpanzee were too similar to distinguish relative divergence of the exons.

Calibration of the Silent Clock. A single-nucleotide change in a codon results in either replacement of the encoded amino acid or the appearance of a synonymous codon. The former component is essentially identical to amino acid replacement in the protein and thus is expected to occur with a regular pace during evolution (protein clock). The latter component also behaves like a clock as demonstrated in sea urchin histone genes (13, 14). The silent clock differs sharply from the protein clock in that it runs with a rapid and even pace for different genes, while in the protein clock the pace is generally slow and differs greatly from protein to protein (12). The rapid rate of the silent clock is particularly suitable for studying evolution of recently diverged species. The second property characteristic of the silent clock—that it runs with an even rate for different genes—allows one to pool data on the genetic distances (K_S^C) obtained from different genes, thus giving statistically reliable estimates for the branching order and timing of species divergence (12).

The genetic distances measured by the number of silent substitutions per site (K_S^C) give the branching pattern of species divergence and the relative times between branching points. To estimate absolute time of divergence, the silent clock must be calibrated from the fossil record. This is equivalent to knowing the evolutionary rate (V_S) of silent

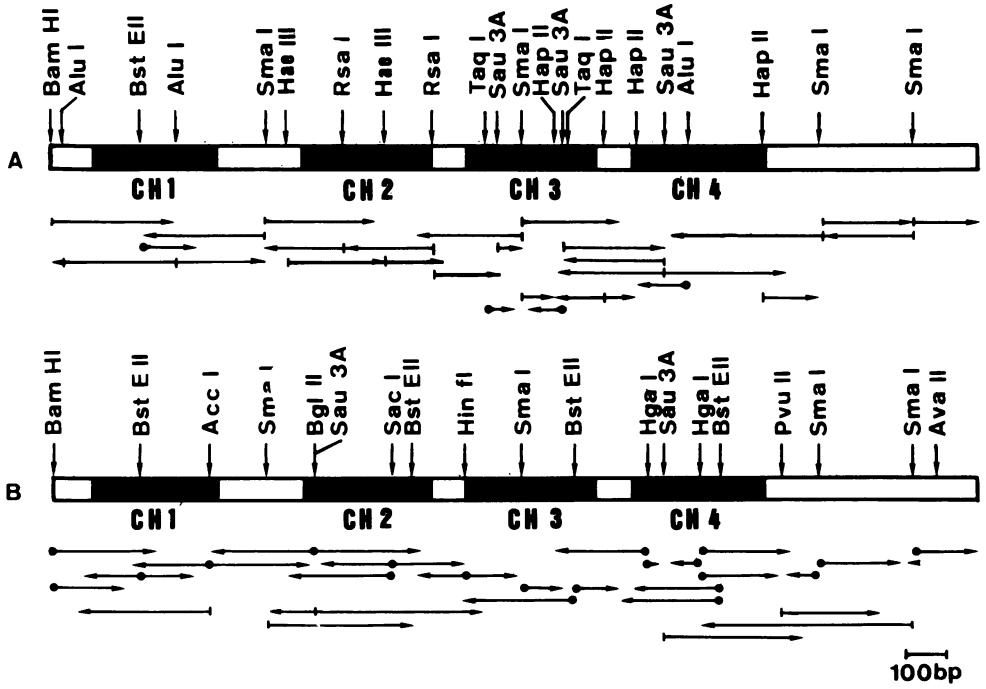

FIG. 1. Sequence strategy and schematic maps for the coding region of the 2.7-kb *Bam*HI fragment (containing $C_{\varepsilon 1}$ gene) of the chimpanzee (*A*) and the orangutan (*B*). Arrows starting at closed circles represent the region determined by the method of Maxam and Gilbert (18). The other arrows indicate regions of the gene sequenced by the dideoxynucleotide chain-termination methods. The four exons are indicated by closed rectangles.

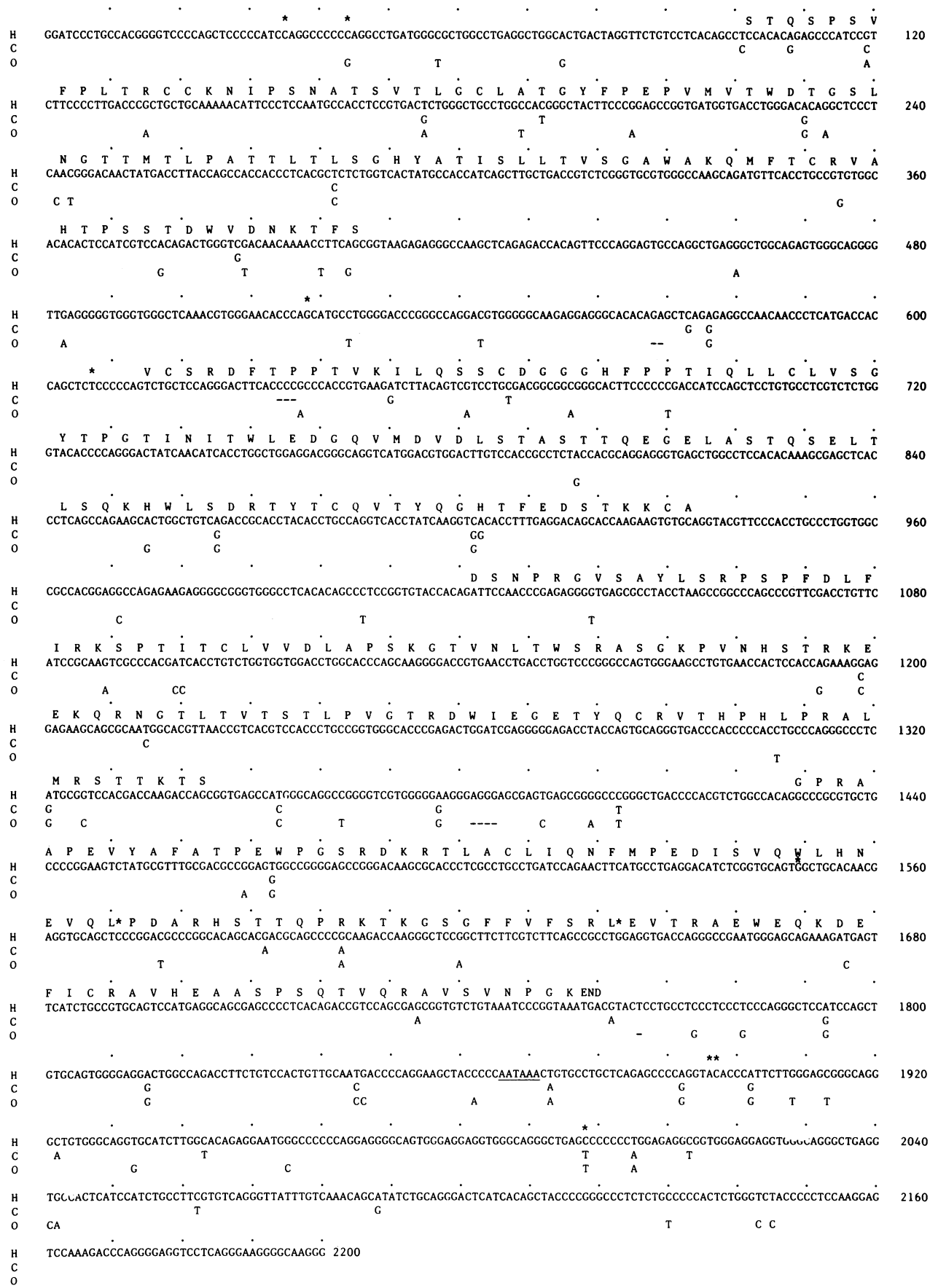

FIG. 2. Nucleotide sequences of the cloned $C_{\varepsilon 1}$ genes from three species: human (H), chimpanzee (C), and orangutan (O). Only nucleotides different from the human sequence are shown. The single-letter amino acid code sequence of the human $C_{\varepsilon 1}$ gene is shown above H. The nucleotide sequence is numbered beginning at the 5′-*Bam*HI sites. We used the nucleotide sequence of the human $C_{\varepsilon 1}$ gene, which was determined in our laboratory (20). Discrepancies between our sequence and the sequence determined by Max *et al.* (21) are indicated by asterisks. The underlined AATAAA represents the presumed poly(A) addition signal.

Table 1. Comparison of the $C_{\varepsilon 1}$ gene sequences of man, chimpanzee, and orangutan for each genetic segment

Segment	Positions	Human vs. chimpanzee			Human vs. orangutan			Chimpanzee vs. orangutan		
		K_A^C	K_S^C	K_N^C	K_A^C	K_S^C	K_N^C	K_A^C	K_S^C	K_N^C
5′-Flanking	1–98			0			0.031			0.031
IVS*1*	408–614			0.010			0.030			0.030
IVS*2*	936–1021			0			0.024			0.024
IVS*3*	1346–1428			0.040			0.093			0.052
3′UT	1761–2200			0.033			0.052			0.047
Noncoding region				0.021			0.045			0.040
C_{H1}	101–407	0.027	0.027		0.045	0.069		0.050	0.083	
C_{H2}	615–935	0.013	0.028		0.017	0.056		0.021	0.042	
C_{H3}	1022–1345	0.008	0.012		0.026	0.037		0.017	0.050	
C_{H4}	1429–1757	0.012	0.015		0.012	0.041		0.008	0.054	
Coding region		0.015	0.020		0.024	0.050		0.023	0.057	

K_A^C and K_S^C, nucleotide differences corrected for superimposed substitutions at the amino acid-changing and -silent positions of protein-encoding region, respectively. K_N^C, nucleotide difference corrected for superimposed substitutions in noncoding region. IVS, intron; 3′UT, 3′-untranslated region; C_H, protein-coding exon.

positions. Nucleotide sequences encoding α_1-antitrypsin from human (25) and baboon (26), β-globin from human (27) and colobus (28), and δ-globin from human (29) and colobus (28) are available at present. From the comparisons between human and Old World monkeys, the mean K_S^C value of the three genes was calculated to be 0.094 (Table 2). Note that as expected, the K_S^C values are similar between different genes in comparison between human and Old World monkeys. Assuming that the separation of hominoids and Old World monkeys occurred 30 million years ago (30), the evolutionary rate of silent substitutions was calculated as $V_S = 0.094/(2 \times 30 \times 10^6) = 1.56 \times 10^{-9}$ substitutions per site per year.

Phylogenetic Tree of Hominoids. To construct the phylogenetic tree among the three species, man, chimpanzee, and orangutan, and to estimate their divergence dates, we used the K_S^C values for the entire coding regions of the $C_{\varepsilon 1}$ genes (Table 2). The K_S^C value for human/chimpanzee versus orangutan was the average of that between human and orangutan, and that between chimpanzee and orangutan. From these K_S^C values, the divergence time (t) can now be calculated as $t = K_S^C/2V_S = K_S^C/(2 \times 1.56 \times 10^{-9})$. The calculated divergence times of chimpanzee and orangutan were 6.4 ± 2.6 and 17.3 ± 4.5 million years, respectively. The estimated mean divergence dates including those of gorilla and gibbon obtained by Sibley and Ahlquist (9) were summarized in Fig. 3. These values are generally consistent with those obtained on the basis of DNA·DNA hybridization kinetics; the datings of the divergence nodes are 6.3–7.7 and 13–16 million years for chimpanzee/human and orangutan, respectively. The estimated divergence time of orangutan is also consistent with the time obtained from the fossil record (30).

During preparation of this manuscript, Koop *et al.* (32) published a comparison of nucleotide sequences of the η-globin pseudogenes of primates and apes. Their results are in good agreement with our study in that orangutan is clearly separated from human, chimpanzee, and gorilla. Although the accurate calibration of evolutionary rate of the pseudogene sequence has not been accomplished, they estimated divergence nodes of the Hominidae (Pongo/Homininae) and the Homininae as 12.7–16.4 million years and 6.3–8.1 million years, respectively. Although the silent clock appears to run with the regular pace at least within lineages including apes and Old World monkeys, the primate genes apparently evolved with a rate slower than those of nonprimate mammals: The mean evolutionary rate of 11 different genes from several mammalian species excluding rodents is estimated to be 3.1×10^{-9} substitutions per site per year by assuming that the date of mammalian divergence is 75 million years ago; rodents were shown to evolve much more rapidly than other mammals (33, 34). This value is larger than the corresponding

Table 2. Evolutionary distances between primates measured by the number of silent substitutions

Species pair	Gene	K_S^C
Human/chimpanzee	$C_{\varepsilon 1}$	0.020 ± 0.008
Human, chimpanzee/orangutan	$C_{\varepsilon 1}$	0.054 ± 0.014
Human/Old World monkeys	α_1-Antitrypsin	0.109 ± 0.020
	β-Globin	0.076 ± 0.037
	δ-Globin	0.096 ± 0.042
	Mean*	0.094 ± 0.027

K_S^C, corrected nucleotide difference per site at the silent position of protein-encoding regions. The standard error was evaluated as K_S^C/N_S, where N_s is the number of sites compared.

*The mean and standard deviation of the K_S^C values of α_1-antitrypsin, β-globin, and δ-globin genes. Nucleotide sequence data sources: α_1-antitrypsin gene from human (25) and baboon (26); β-globin gene from human (27) and colobus (28); δ-globin gene from human (29) and colobus (28); $C_{\varepsilon 1}$ gene from human (20).

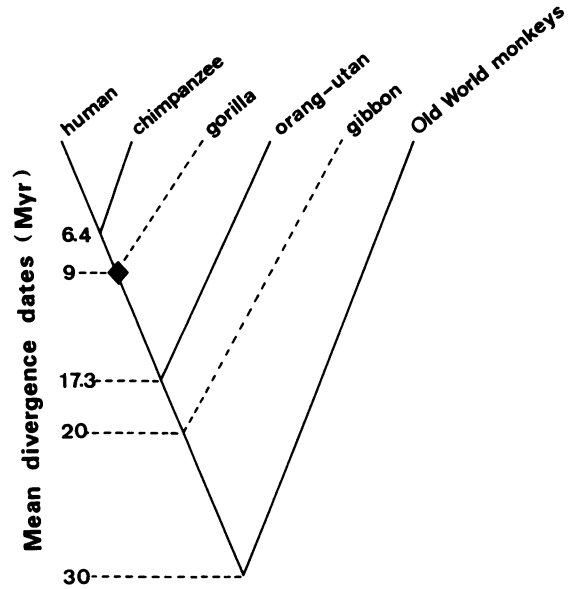

FIG. 3. Phylogenetic tree of primates and mean divergence dates. Data for the dates of divergence nodes for gorilla and gibbon were taken from Sibley and Ahlquist (9). ◆, An immediate ancestor of the African apes and man or an African ape with similarity to gorilla, which is placed at ≈9 million years ago (30, 31).

value of the primate genes. The decelerated rate of hominoid evolution was also observed in the η-globin pseudogene (35). The evolutionary clock of nonprimate genes leads to an unreasonable estimate for the divergence of apes and Old World monkeys; their separation was estimated to have occurred about 15 million years ago. It may be of interest to know in what lineage the evolutionary rate of primate genes has slowed down. Sequence data from New World monkeys and prosimians would clarify this point.

Evolution of Immunoglobulin ε Genes in Primates. The human C_ε gene family comprises one active ($C_{\varepsilon 1}$) gene and two pseudogenes ($C_{\varepsilon 2}$ and $C_{\varepsilon 3}$). The $C_{\varepsilon 1}$ gene and $C_{\varepsilon 2}$ truncated pseudogene are linked in the human C_H gene locus on chromosome 14, whereas the $C_{\varepsilon 3}$ processed pseudogene is translocated to chromosome 9 (36). The $C_{\varepsilon 2}$ pseudogene was present in man and gorilla, whereas it was absent in chimpanzee when examined by Southern blot hybridization (11). The $C_{\varepsilon 3}$ pseudogene is found in all the apes and Old World monkeys examined so far (11). The results suggested two major alternative possibilities: (*i*) gorilla is closer to man than chimpanzee is or (*ii*) chimpanzee has lost the $C_{\varepsilon 2}$ gene after its divergence. Cloning and characterization of DNA segments surrounding the C_ε and C_α genes of chimpanzee and gorilla showed that chimpanzee had lost the entire exons of the $C_{\varepsilon 2}$ gene (unpublished data). The results are consistent with the present phylogenetic tree.

We are grateful to Dr. O. Takenaka for providing primate blood samples. We thank Ms. S. Okamoto and K. Mitsuyasu for preparation of this manuscript. This investigation was supported in part by grants from the Ministry of Education, Science and Culture of Japan.

1. Zuckerkandl, E. & Pauling, L. (1962) in *Horizons in Biochemistry*, eds. Kasha, M. & Pullman, B. (Academic, New York), pp. 189–225.
2. Zuckerkandl, E. & Pauling, L. (1965) in *Evolving Genes and Proteins*, eds. Brysonm, V. & Vogel, H. (Academic, New York), pp. 97–166.
3. Dickerson, R. E. (1971) *J. Mol. Evol.* **1**, 26–45.
4. Wilson, A. C., Carlson, S. S. & White, T. J. (1977) *Annu. Rev. Biochem.* **46**, 573–639.
5. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, England), pp. 55–97.
6. Sarich, V. M. & Wilson, A. C. (1976) *Proc. Natl. Acad. Sci. USA* **58**, 142–148.
7. Sarich, V. M. & Wilson, A. C. (1976) *Science* **158**, 1200–1203.
8. Andrews, P. & Cronin, J. E. (1982) *Nature (London)* **297**, 541–546.
9. Sibley, C. G. & Ahlquist, J. E. (1984) *J. Mol. Evol.* **20**, 2–15.
10. Brown, W. M., George, M., Jr. & Wilson, A. C. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 1967–1971.
11. Ueda, S., Takenaka, O. & Honjo, T. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3712–3715.
12. Miyata, T., Yasunaga, T. & Nishida, T. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 7328–7332.
13. Busslinger, M., Rusconi, S. & Birnstiel, M. L. (1982) *EMBO J.* **1**, 27–33.
14. Miyata, T., Hayashida, H., Kikuno, R., Toh, H. & Kawade, Y. (1985) in *Evolution of Interferon Genes*, ed. Grosser, I. (Academic, London), pp. 1–30.
15. Brown, W. M., Prager, E. M., Wang, A. & Wilson, A. C. (1982) *J. Mol. Evol.* **18**, 225–239.
16. Hasegawa, M., Yano, T. & Kishino, H. (1984) *Proc. Jpn. Acad. Ser. B* **60**, 95–98.
17. Nikaido, T., Yamawaki-Kataoka, Y. & Honjo, T. (1982) *J. Biol. Chem.* **257**, 7322–7329.
18. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
19. Messing, J. (1983) *Methods Enzymol.* **101**, 20–78.
20. Ueda, S., Nakai, S., Nishida, Y., Hisajima, H. & Honjo, T. (1982) *EMBO J.* **1**, 1539–1544.
21. Max, E. E., Battey, J., Ney, R., Kirsch, R. I. & Leder, P. (1982) *Cell* **29**, 691–699.
22. Miyata, T. & Yasunaga, T. (1980) *J. Mol. Evol.* **16**, 23–26.
23. Jukes, T. H. & Cantor, C. R. (1969) in *Mammalian Protein Molecules*, ed. Munro, H. N. (Academic, New York), Vol. 3, pp. 21–132.
24. Nishida, Y., Ueda, S., Takahashi, N., Nakai, S., Hisajima, H., Barker, P. E., Ruddle, F. H. & Honjo, T. (1983) in *Humoral Factors in Host Defense*, eds. Yamamura, Y., Yahashi, H., Honjo, T., Kishimoto, T., Muramatsu, M. & Osawa, T. (Academic, New York), pp. 323–336.
25. Rosenberg, S., Barr, P. J., Najarian, R. C. & Hallewell, A., Jr. (1984) *Nature (London)* **312**, 77–80.
26. Kurachi, K., Chandra, T., Frienzner Degen, S. J., White, T. T., Marchioro, T. L., Woo, S. L. C. & Davie, E. W. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6826–6830.
27. Lawn, R. M., Efstratiadis, A., O'Connell, C. & Maniatis, T. (1980) *Cell* **21**, 647–651.
28. Martin, S. L., Vincent, K. A. & Wilson, A. C. (1983) *J. Mol. Biol.* **164**, 513–528.
29. Spritz, R. A., DeRiel, J. K., Forget, B. G. & Weissman, S. M. (1980) *Cell* **21**, 639–646.
30. Andrews, P. (1985) *Nature (London)* **314**, 498–499.
31. Ishida, H., Pickford, M., Nakaya, H. & Nakano, Y. (1984) *Afr. Stud. Monog.*, Suppl. 2, 73–85.
32. Koop, B. F., Goodman, M., Xu, P., Chan, K. & Slightom, J. L. (1986) *Nature (London)* **319**, 234–238.
33. Kikuno, R., Hayashida, H. & Miyata, T. (1985) *Proc. Jpn. Acad. Ser. B* **61**, 153–156.
34. Wu, C.-I., & Li, W.-H. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1741–1745.
35. Goodman, M., Koop, B. F., Czelusniak, J. & Weiss, M. L. (1984) *J. Mol. Biol.* **180**, 803–823.
36. Battey, J., Max, E. E., McBride, W. O., Swan, D. & Leder, P. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5956–5960.