Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 3712–3715, June 1985
Evolution

A truncated immunoglobulin ϵ pseudogene is found in gorilla and man but not in chimpanzee

(qualitative molecular evolution/human evolution)

SHINTAROH UEDA*, OSAMU TAKENAKA†, AND TASUKU HONJO‡§

*Department of Anthropology, Faculty of Science, The University of Tokyo, Tokyo 113, Japan; †Department of Biochemistry, Primate Research Institute, Kyoto University, Aichi 484, Japan; and ‡Department of Genetics, Osaka University Medical School, Osaka 530, Japan

Communicated by Motoo Kimura, January 22, 1985

ABSTRACT **Molecular genetic analyses of the young pseudogenes of the immunoglobulin C_ϵ genes were carried out to obtain qualitative evidence for the phylogenetic branching pattern of hominoid primates. We found that Old World monkeys had two C_ϵ genes, one of which was processed. Among the hominoids examined only the gorilla and human genomes contained three C_ϵ genes: an active, a truncated, and a processed gene. Other hominoids so far examined, including chimpanzee, contained two C_ϵ genes: one active and the other processed. These results suggest that the processed C_ϵ pseudogene was generated before the divergence between Old World monkeys and hominoids and that the gorilla is more closely related to man than the chimpanzee is, unless the chimpanzee has lost the $C_\epsilon 2$ gene after the divergence of this species.**

The immunoglobulin genes in the heavy-chain constant (C_H) region cluster are divided into five classes, C_μ, C_δ, C_γ, C_ϵ, and C_α. There are three C_ϵ genes in the human genome (1–3). We have called the C_ϵ genes in the 2.7-, 5.9-, and 8.0-kilobase (kb) *Bam*HI fragments the $C_\epsilon 1$, $C_\epsilon 2$, and $C_\epsilon 3$ genes, respectively (2, 4, 5). One of them ($C_\epsilon 1$) is active, whereas the remaining two are pseudogenes. One pseudogene ($C_\epsilon 2$) is truncated by recombination (3, 4) and the other ($C_\epsilon 3$) is processed (5, 6). The latter lacks introns entirely and is translocated from chromosome 14 to 9, suggesting that this gene was created by reverse transcription of an aberrantly transcribed C_ϵ sequence. Since the C_H gene family of mouse contains only one C_ϵ gene (7), the creation of two C_ϵ pseudogenes in the human genome seems to have taken place after mammalian radiation. Comparison of the nucleotide sequences (4, 5) allowed us to estimate that the $C_\epsilon 2$ and $C_\epsilon 3$ genes diverged from the $C_\epsilon 1$ gene $6.6–8.9 \times 10^6$ and 39×10^6 years ago, respectively.

The primate superfamily Hominoidea includes man, the chimpanzee, the pygmy chimpanzee, the gorilla, the orangutan, and the gibbons. The branching sequence of the lineages and the datings of the divergence nodes are still in dispute. A large number of studies on hominoid relationships have been based on morphology, fossils, behavior, and molecular comparison. Usually the most powerful evidence for the study of phylogeny is derived from fossil records and molecular comparisons, both of which have, unfortunately, limitations for studies on hominoid evolution. The fossil records of hominoids, especially nonhuman hominoids, are too scarce to draw a definitive conclusion. Comparison of amino acid and nucleotide sequences of primates is not convincing either because the divergence of the sequences is too small to quantitate accurately a small difference in the branching time.

In this report we present another strategy to determine the branching sequence of the lineages: using human C_ϵ pseudogenes, which evolved very recently. The presence or the absence of young pseudogenes in DNA of various species is able to provide a qualitative answer to determine the branching sequence. We studied the organization of the C_ϵ genes of nonhuman primates, including 13 species of Old World monkeys and 5 species of hominoids, to analyze their evolutionary relationships to man.

MATERIALS AND METHODS

The species of nonhuman primates examined include 13 species of Cercopithecoidea (Old World monkeys): *Macaca fuscata* (Japanese monkey), *Macaca mulatta* (rhesus monkey), *Macaca fascicularis* (crab-eating monkey), *Macaca arctoides* (red-faced macaque), *Macaca nemestrina* (pig-tailed macaque), *Macaca cyclopis* (Formosan monkey), *Macaca radiata* (bonnet monkey), *Macaca assamensis* (Assamese monkey), *Theropithecus gelada* (gelada), *Papio anubis* (anubis baboon), *Papio hamadryas* (hamadryas baboon), *Erythrocebus patas* (patas monkey), and *Cercopithecus aethiops* (green monkey); and 5 species of Hominoidea (hominoids): *Pan troglodytes* (chimpanzee), *Gorilla gorilla* (gorilla), *Pongo pygmaeus* (orangutan), *Hylobates lar* (white-handed gibbon), and *Hylobates agilis* (agile gibbon).

DNA was prepared from lymphocytes of peripheral blood (8), except for DNAs of orangutan and gorilla. DNA of orangutan was prepared not only from lymphocytes but also from an Epstein–Barr-virus-transformed cell line of the same individual. DNA of gorilla was obtained from lymph nodes and further purified by using equilibrium sedimentation in a cesium chloride density gradient. Two micrograms of DNA was digested with appropriate amounts of restriction enzymes, electrophoresed in 0.5% agarose gels, and transferred to nitrocellulose filters (9). DNA fragments used as probes were labeled with [α-^{32}P]dCTP by nick-translation to a specific activity of 500–1000 cpm/pg of DNA (10). Hybridization was carried out in 1 M NaCl at 65°C as described previously (11) and filters were washed three times (30 min each) in 0.15 M NaCl/0.015 M sodium citrate/0.1% sodium dodecyl sulfate at 65°C.

RESULTS

The organization of C_ϵ genes in DNAs from various species of primates was studied by the Southern hybridization method, using human C_ϵ probes. The number of C_ϵ genes in the genome of each species was estimated by using a 1.2-kb *Bam*HI/*Hpa* I fragment of the human $C_\epsilon 1$ gene as probe (probe A shown in Fig. 1), while the processed C_ϵ gene was analyzed by using a 1.0-kb *Acc* I/*Acc* I fragment of the human $C_\epsilon 3$ gene as probe (probe B shown in Fig. 1). Probe A cross-hybridized with the other C_ϵ genes, while probe B was specific for the processed C_ϵ gene under the stringent washing

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: kb, kilobase.

§Present address: Department of Medical Chemistry, Faculty of Medicine, Kyoto University, Kyoto 606, Japan.

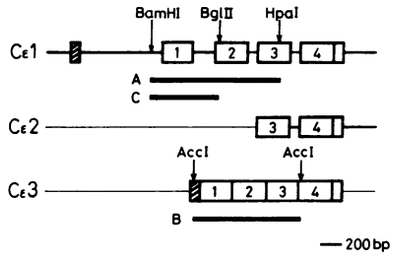

FIG. 1. Structure of the human $C_\epsilon 1$, $C_\epsilon 2$, and $C_\epsilon 3$ genes. The C_H exons are shown by open boxes with domain numbers. The 5′ flanking region homologous in the $C_\epsilon 1$ and $C_\epsilon 3$ genes and the 3′ untranslated sequence are shown by hatched and open boxes, respectively. The regions homologous to the $C_\epsilon 1$ gene in the $C_\epsilon 2$ and $C_\epsilon 3$ genes are shown by rectangles, while nonhomologous regions are shown by lines. Fragments (A, B, and C) used as probes are indicated by horizontal bars. bp, Base pairs.

conditions, although another, very weak, band corresponding to the $C_\epsilon 1$ gene sometimes appeared under the mild washing conditions.

Probe A detected two hybridizing bands in *Bam*HI digests of all the Old World monkey DNAs examined (Fig. 2*A*): 2.7- and 7.1-kb bands in all the macaques, baboons, and gelada; 7.1- and 7.4-kb bands in the patas monkey; and 7.1- and 7.6-kb bands in the green monkey. Probe B detected only 7.1-kb *Bam*HI bands in all the Old World monkey DNAs (Fig. 2*A*). As shown in Fig. 2*B*, *Bam*HI digests of gibbon, orangutan, and chimpanzee DNAs contained two bands, each hybridizing with probe A. *Bam*HI bands other than 2.7 kb were also detectable with probe B (data not shown). We examined 11 individual chimpanzee DNAs, but there was no variation in hybridization patterns.

In contrast to the hominoid DNAs mentioned above, there were three bands (2.7, 6.9, and 15 kb) hybridizing with probe A in *Bam*HI digests of the gorilla DNA (Fig. 2*B*). When probe B was used, the 15-kb *Bam*HI band was shown to contain the processed C_ϵ gene (data not shown). So we examined whether gorilla DNA contained a truncated C_ϵ ($C_\epsilon 2$) gene as in the human genome. To identify the $C_\epsilon 2$ gene, which lacks the $C_H 1$ and $C_H 2$ exons in addition to their flanking sequences (3, 4), we have used as probe the 0.65-kb *Bam*HI/*Bgl* II human $C_\epsilon 1$ gene fragment (probe C shown in Fig. 1), which is deleted in the human $C_\epsilon 2$ gene. Since probe C cross-hybridizes with the $C_\epsilon 3$ gene, which contains the pseudo-$C_H 1$ and pseudo-$C_H 2$ exons, the C_ϵ gene fragments that do not hybridize with probe C are the truncated $C_\epsilon 2$ gene fragments. These are the 5.9-kb *Bam*HI band of the human DNA and the 6.9-kb *Bam*HI band of the gorilla DNA (Fig. 2*B*). In DNAs of other hominoids examined, probe C hybridized to all the fragments hybridizing to probe A (data not shown). These results suggest that there are three C_ϵ genes in the gorilla genome, like the human genome: an active, a processed, and a truncated gene.

Using several other restriction enzymes, we confirmed that Old World monkeys and hominoids except for the gorilla and man contained two C_ϵ genes. Results of *Eco*RI digestions are shown in Fig. 2*C*. *Eco*RI digests of orangutan and chimpanzee DNAs produced two bands each hybridizing with probe A (lanes 2 and 3). Although *Eco*RI digests of the white-handed gibbon and agile gibbon DNAs yielded three bands (lanes 1 and 4), another individual DNA of the white-handed gibbon revealed two *Eco*RI bands (lane 5). Probe B detected one band in all the digests except for the gibbon samples having three *Eco*RI bands, in which both of the smaller two bands hybridized with probe B (data not shown), suggesting the presence of the *Eco*RI restriction fragment length polymorphism of the $C_\epsilon 3$ processed gene in the gibbon.

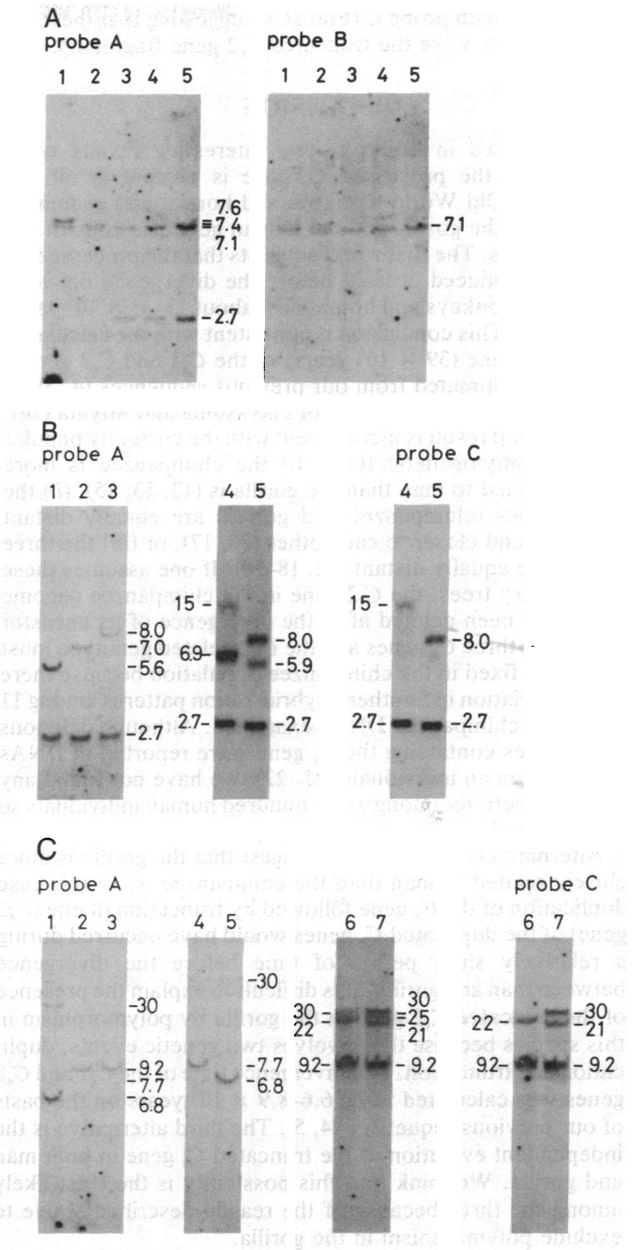

FIG. 2. Southern hybridization of primate DNAs with human C_ϵ probes. Numbers indicate sizes of the hybridizing fragments in kb. (*A*) *Bam*HI digests of Old World monkey DNAs with probe A and probe B. The DNA source of each lane is as follows: 1, green monkey; 2, patas monkey; 3, hamadryas baboon; 4, gelada; and 5, Japanese monkey. (*B*) *Bam*HI digests of various hominoid DNAs with probe A and probe C. The DNA source of each lane is as follows: 1, white-handed gibbon; 2, orangutan; 3, chimpanzee; 4, gorilla; and 5, man. (*C*) *Eco*RI digests of various hominoid DNAs with probe A and probe C. The DNA source of each lane is as follows: 1 and 5, white-handed gibbon; 2, orangutan; 3, chimpanzee; 4, agile gibbon; 6, gorilla; and 7, man.

There were three bands (9.2, 22, and 30 kb) hybridizing with probe A in *Eco*RI digests of the gorilla DNA (lane 6). By using probe B the 9.2-kb *Eco*RI band was shown to contain the processed C_ϵ gene (data not shown). The three *Eco*RI bands, 9.2, 25, and 30 kb, seen most frequently in the human genome contain the $C_\epsilon 3$, $C_\epsilon 2$, and $C_\epsilon 1$ genes, respectively. The 21-kb *Eco*RI band in the human DNA sample shown in Fig. 2*C* is due to the restriction fragment length polymorphism of the $C_\epsilon 1$ gene. The 25-kb *Eco*RI band of the human DNA and the 30-kb *Eco*RI band of the gorilla DNA were

undetectable with probe C (Fig. 2*C*), suggesting that these C_ϵ gene fragments were the truncated $C_\epsilon 2$ gene fragments.

DISCUSSION

As summarized in Table 1, two interesting results were obtained: (*i*) the processed C_ϵ gene is present in all the catarrhines (Old World monkeys and hominoids) examined and (*ii*) only the gorilla and the human genomes contain the three C_ϵ genes. The first result suggests that the processed C_ϵ gene was produced at least before the divergence between Old World monkeys and hominoids, about 13–33 $\times 10^6$ years ago (12, 13). This conclusion is consistent with the calculated divergence time (39×10^6 years) of the $C_\epsilon 1$ and $C_\epsilon 3$ genes, which was estimated from our previous sequences (4, 5) by a slightly modified procedure of Hayashida and Miyata (14).

The second result is inconsistent with the currently popular hominoid phylogenetic trees: (*i*) the chimpanzee is more closely related to man than the gorilla is (12, 13, 15), (*ii*) the African apes (chimpanzee and gorilla) are equally distant from man and closer to each other (16, 17), or (*iii*) the three species are equally distant (12, 18–20). If one assumes these evolutionary trees, the $C_\epsilon 2$ gene in the chimpanzee genome must have been deleted after the divergence of its ancestor containing three C_ϵ genes and the $C_\epsilon 2$-deleted genotype must have been fixed in the chimpanzee population because there was no variation in Southern hybridization patterns among 11 individual chimpanzee DNAs examined. Although deletions of C_H genes containing the C_ϵ gene were reported in DNAs of a few human individuals (21, 22), we have not found any C_ϵ gene deletions among a few hundred human individuals so far examined.

Alternatively, our findings suggest that the gorilla is more closely related to man than the chimpanzee is. In this case duplication of the C_ϵ gene followed by truncation in one ($C_\epsilon 2$ gene) of the duplicated C_ϵ genes would have occurred during a relatively short period of time before the divergence between man and gorilla. It is difficult to explain the presence of the truncated $C_\epsilon 2$ gene in the gorilla by polymorphism in this species because this involves two genetic events, duplication and truncation. The divergence time of the $C_\epsilon 1$ and $C_\epsilon 2$ genes was calculated to be 6.6–8.9 $\times 10^6$ years on the basis of our previous sequences (4, 5). The third alternative is the independent evolution of the truncated C_ϵ gene in both man and gorilla. We think that this possibility is the least likely among the three because of the reason described above to exclude polymorphism in the gorilla.

The phylogenetic branching pattern and the dating of divergence nodes of man, chimpanzee, and gorilla are still being debated. In all the studies so far available the divergence time was calculated by calibration of measured differences in morphological characters (16), chromosomal banding pattern (15), amino acid sequences (12), DNA-DNA reassociation kinetics (13), and restriction endonuclease cleavage maps and nucleotide sequences of mitochondrial DNA (18–20) against an external dating source such as fossils or geological events. Obviously these studies include several assumptions for calculation. The present study provides molecular genetic evidence that qualitatively distinguishes chimpanzee from man and gorilla. Since this type of study gives a yes-or-no answer without any assumption, extensive analyses of many other young pseudogenes in primate DNAs will allow us to construct a reliable phylogenetic tree of primates. These studies will open an approach to the study of molecular evolution, which may be called qualitative molecular evolutionary analysis.

Table 1. Organization of the C_ϵ genes in primates

Species	*Bam*HI fragment size, kb		
	$C_\epsilon 1$ gene (active)	$C_\epsilon 2$ gene (truncated)	$C_\epsilon 3$ gene (processed)
Man	2.7	5.9	8.0
Gorilla	2.7	6.9	15
Chimpanzee	2.7	—	8.0
Orangutan	2.7	—	7.0
Gibbons	2.7	—	5.6
Macaques	2.7	—	7.1
Baboons	2.7	—	7.1
Gelada	2.7	—	7.1
Patas monkey	7.4	—	7.1
Green monkey	7.6	—	7.1

In addition to the three C_ϵ genes there are five C_γ and two C_α genes in the human immunoglobulin genes. The order of human C_H genes is proposed to be 5′-C_μ–C_δ–$C_\gamma 3$–$C_\gamma 1$–$C_\epsilon 2$–$C_\alpha 1$–ψC_γ–$C_\gamma 2$–$C_\gamma 4$–$C_\epsilon 1$–$C_\alpha 2$-3′ (3, 4, 21–28), indicating the duplication of the set of the C_H genes involving the C_γ, C_ϵ, and C_α genes. The relationship between man and the African apes can be further tested by studying the organization of the C_H genes. A high degree of length polymorphism in the C_γ genes (21, 28) and the absence of appropriate restriction enzymes suitable for estimation of the C_α gene number in nonhuman primates' genomes prevented us from estimation of the correct numbers of the C_γ and C_α genes in the gorilla genome by the Southern hybridization method alone. Recent cloning of two sets of the C_H genes involving the C_ϵ and C_α genes from the gorilla genome (unpublished) suggests that the gorilla has a C_H gene organization similar to that of man and makes it more difficult to explain the presence of the $C_\epsilon 2$ gene in the gorilla genome by polymorphism or independent evolution.

We thank T. Ishida for establishment of the Epstein–Barr virus-transformed cell line of orangutan and providing us with the cells, Dr. H. Yoshizawa (the Tokyo Metropolitan Institute of Medical Science) for providing us with chimpanzee blood samples, and Prof. K. Omoto for his encouragement. We are grateful to Drs. H. Hayashida and T. Miyata of Kyushu University for calculation of the C_ϵ pseudogene divergence time. We also thank Dr. R. Kominami for his critical reading of this manuscript. This study was supported in part by grants from the Ministry of Education, Science and Culture and by a cooperative research grant from Primate Research Institute, Kyoto University.

1. Flanagan, J. G. & Rabbitts, T. H. (1982) *EMBO J.* **1**, 655–660.
2. Nishida, Y., Miki, T., Hisajima, H. & Honjo, T. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 3833–3837.
3. Max, E. E., Battey, J., Ney, R., Kirsch, I. R. & Leder, P. (1982) *Cell* **29**, 691–699.
4. Hisajima, H., Nishida, Y., Nakai, S., Takahashi, N., Ueda, S. & Honjo, T. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2995–2999.
5. Ueda, S., Nakai, S., Nishida, Y., Hisajima, H. & Honjo, T. (1982) *EMBO J.* **1**, 1539–1544.
6. Battey, J., Max, E. E., McBride, W. O., Swan, D. & Leder, P. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5956–5960.
7. Shimizu, A., Takahashi, N., Yamawaki-Kataoka, Y., Nishida, Y., Kataoka, T. & Honjo, T. (1981) *Nature (London)* **289**, 149–153.
8. Bell, G. I., Karam, J. H. & Rutter, W. J. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 5759–5763.
9. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
10. Maniatis, T., Jeffrey, A. & Kleid, D. G. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1184–1188.
11. Honjo, T., Obata, M., Yamawaki-Kataoka, Y., Kataoka, T., Kawakami, T., Takahashi, N. & Mano, Y. (1979) *Cell* **18**, 559–568.
12. Goodman, M., Braunitzer, G., Stangl, A. & Schrank, B. (1983) *Nature (London)* **303**, 546–548.
13. Sibley, C. G. & Ahlquist, J. E. (1984) *J. Mol. Evol.* **20**, 2–15.

14. Hayashida, H. & Miyata, T. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 2671–2675.
15. Yunis, J. J. & Prakasho, O. (1982) *Science* **215,** 1525–1530.
16. Kluge, A. G. (1983) in *New Interpretations of Ape and Human Ancestry*, eds. Ciochon, R. L. & Corruccini, R. S. (Plenum, New York), pp. 151–177.
17. Templeton, A. R. (1983) *Evolution* **37,** 221–244.
18. Brown, W. M., Prager, E. M., Wang, A. & Wilson, A. C. (1982) *J. Mol. Evol.* **18,** 225–239.
19. Brown, W. M. (1983) in *Evolution of Genes and Proteins*, eds. Nei, M. & Koehn, R. K. (Sinauer Assoc., Sunderland, MA), pp. 62–88.
20. Ferris, S. D., Wilson, A. C. & Brown, W. M. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 2432–2436.
21. Lefranc, M., Lefranc, G. & Rabbitts, T. H. (1982) *Nature (London)* **300,** 760–762.
22. Migone, N., Oliviero, S., de Lange, G., Delacroix, D. L., Boschis, D., Altruda, F., Silengo, L., DeMarchi, M. & Carbonara, A. O. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 5811–5815.
23. Takahashi, N., Ueda, S., Obata, M., Nikaido, T., Nakai, S. & Honjo, T. (1982) *Cell* **29,** 671–679.
24. Krawinkel, U. & Rabbitts, T. H. (1982) *EMBO J.* **1,** 403–407.
25. Ellison, J. & Hood, L. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 1984–1988.
26. Flanagan, J. G. & Rabbitts, T. H. (1982) *Nature (London)* **300,** 709–713.
27. Flanagan, J. G., Lefranc, M. & Rabbitts, T. H. (1984) *Cell* **36,** 681–688.
28. Bech-Hansen, N. T., Linsley, P. S. & Cox, D. W. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 6952–6956.