J. Mol. Biol. (1989) **205**, 85–90

Nucleotide Sequences of Immunoglobulin-Epsilon Pseudogenes in Man and Apes and Their Phylogenetic Relationships

Shintaroh Ueda[1], Yoshihisa Watanabe[1], Naruya Saitou[1] Keiichi Omoto[1], Hidenori Hayashida[2]†, Takashi Miyata[2] Hiroshi Hisajima[3]‡ and Tasuku Honjo[3]

[1]*Department of Anthropology, Faculty of Science The University of Tokyo, Tokyo 113, Japan*

[2]*Department of Biology, Faculty of Science Kyushu University, Fukuoka 812, Japan*

[3]*Department of Medical Chemistry Kyoto University Faculty of Medicine Kyoto 606, Japan*

(Received 13 May 1988, and in revised form 12 August 1988)

To understand the phylogenetic relationships between hominoids, the nucleotide sequences of immunoglobulin-epsilon processed pseudogenes from chimpanzee, gorilla and orangutan were determined. The basic structures of these processed pseudogenes agreed with their human counterpart. Although the degrees of nucleotide differences between man and the African apes had no statistical significance, all the analytical data examined supported the theory that chimpanzee is the closest relative of man. This result was consistent with that deduced by our recent qualitative study.

Studies on the nucleotide sequences of globin genes have suggested that the molecular clock runs more slowly in hominoids than in non-hominoid primates. According to the present data, however, further retardation of the evolutionary rate was not observed in the human lineage. Assuming that orangutan diverged 14 million years ago and that the evolutionary rate between the orangutan lineage and the lineage leading to the other three species is constant, the divergence dates of chimpanzee and gorilla were estimated to be 4·9(±0·9) and 5·9(±0·9) million years ago, respectively.

1. Introduction

What is the closest relative of man among the living apes? A large number of molecular studies have been carried out to elucidate the phylogenetic relationships between man and apes and also their divergence dates. It is now popularly accepted that the orangutan is clearly separated from man and the African apes (chimpanzee and gorilla). There is, however, no good agreement about the order of branching nodes between man and the African apes (Koop *et al.*, 1986; Sibley & Ahlquist, 1987, and references therein). This is because the three species are so closely related that the divergence of the nucleotide sequences of most genes, or the amino acid sequences of their products, between these species is too small to evaluate with a statistical significance. To overcome this problem there are at least two approaches: the quantitative accumulation of a large number of sequence data and the qualitative distinction of the gene organization (Ueda *et al.*, 1985, 1988). Studies on pseudogenes have an advantage over those on functional genes for the former strategy in that they have accumulated more nucleotide changes. We have determined and compared the nucleotide sequences of the functional immunoglobulin C_ε genes of man, chimpanzee and orangutan, and estimated the divergence date between man and chimpanzee to be about 6·4 million years ago (Sakoyama *et al.*, 1987).

† Present address: National Institute of Genetics, Mishima 411, Japan.

‡ Present address: Fujisawa Pharmaceutical Co. Ltd, Tsukuba Exploratory Research Laboratories, Ibaragi 300-26, Japan.

0022-2836/89/010085-06 $03.00/0

We reported a preliminary comparison of nucleotide sequences of the immunoglobulin $C_{\varepsilon 3}$ pseudogene (a processed gene) between man, chimpanzee and gorilla and concluded that chimpanzee is closer than gorilla to man (Ueda *et al.*, 1982, 1986).

In this report we present the nucleotide sequences of the immunoglobulin $C_{\varepsilon 3}$ genes of chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*) and orangutan (*Pongo pygmaeus*) and compare them with their human counterpart (Ueda *et al.*, 1982). On the basis of this comparison, we also estimate the divergence dates of the African apes.

2. Materials and Methods

Restriction endonucleases (*Alu*I, *Bam*HI, *Eco*RI, *Hae*III, *Hap*II, *Pst*I, *Pvu*II, *Rsa*I and *Sau*3AI), phage T4 DNA ligase, bacterial alkaline phosphatase and M13 sequencing kit were purchased from Takara Shuzo Co. Ltd (Kyoto) or Toyobo Co. Ltd (Osaka). [α-^{32}P]dCTP (approx. 3000 Ci/mmol) was from New England Nuclear. High molecular weight DNAs of peripheral blood of chimpanzee and orangutan, and lymph nodes of gorilla were prepared as described (Ueda *et al.*, 1985). The $C_{\varepsilon 3}$ genes of chimpanzee, gorilla and orangutan are found in 8·0, 15 and 7·0 kb *Bam*HI fragments, respectively (Ueda *et al.*, 1985). Each high molecular weight DNA was completely digested with *Bam*HI and fractionated by electrophoresis in 0·5% (w/v) agarose gel. These *Bam*HI fragments were cloned into Charon 28 (Rimm *et al.*, 1980) or its derivative phage, designated as Charon 28HS (Sakoyama *et al.*, 1987). These recombinant phage DNAs were packaged into coat proteins *in vitro*. Phage plaques were screened using the human $C_{\varepsilon 1}$ and $C_{\varepsilon 3}$ genes (Ueda *et al.*, 1982) as probes. The nucleotide sequences were determined by the dideoxynucleotide chain termination method using M13mp10 and M13mp11 (Messing, 1983; Norrander *et al.*, 1983; Sanger *et al.*, 1977).

3. Results and Discussion

(a) *The basic structures of the $C_{\varepsilon 3}$ genes of chimpanzee, gorilla and orangutan*

There are two or three *Bam*HI fragments that hybridize with the human $C_{\varepsilon 1}$ gene as probe in the man, chimpanzee, gorilla and orangutan genomes; 2·7 kb, 5·9 kb and 8·0 kb in man; 2·7 kb and 8·1 kb in chimpanzee; 2·7 kb, 6·9 kb and 15 kb in gorilla; 2·7 kb and 7·0 kb in orangutan (Nishida *et al.*, 1982; Ueda *et al.*, 1985, 1986). Whereas all the 2·7 kb bands are the active $C_{\varepsilon 1}$ gene fragments, the 5·9 kb (man) and 6·9 kb (gorilla) bands are the $C_{\varepsilon 2}$ pseudogene that is truncated by recombination. The remaining bands, 8·0 kb (man), 8·0 kb (chimpanzee), 15 kb (gorilla) and 7·0 kb (orangutan), have been shown to be the $C_{\varepsilon 3}$ processed pseudogene (Ueda *et al.*, 1985, 1986). Figure 1 shows the restriction endonuclease maps of the cloned DNA fragments containing the $C_{\varepsilon 3}$ genes.

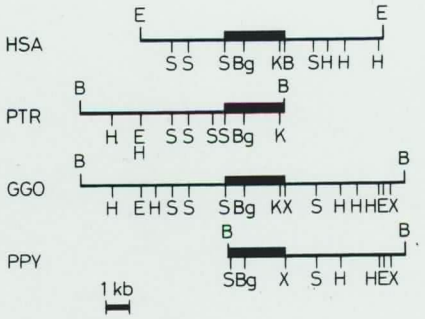

Figure 1. Aligned restriction maps of the human (HSA), chimpanzee (PTR), gorilla (GGO) and orangutan (PPY) $C_{\varepsilon 3}$ genes. The filled boxes indicate the regions containing the $C_{\varepsilon 3}$ genes. Restriction enzymes are abbreviated as follows; B: *Bam*HI, Bg: *Bgl*II, E: *Eco*RI, H: *Hin*dIII, K: *Kpn*I, S: *Sac*I and X: *Xba*I.

The nucleotide sequences of the $C_{\varepsilon 3}$ genes of chimpanzee, gorilla and orangutan were determined and compared with that of the human $C_{\varepsilon 3}$ gene as shown in Figure 2. The basic structures of these $C_{\varepsilon 3}$ genes are similar: the $C_{\varepsilon 3}$ genes lack the three introns entirely at the splicing signals and have A-rich sequences 16 bp 3′ to the putative poly(A) addition signal. Moreover, the 100 bp segment homologous to the 5′ flanking region of the $C_{\varepsilon 1}$ gene (active gene), which is located 754 bp upstream from the $C_{\varepsilon 1}$ coding region, is spliced to the pseudo-CH1 exon of the $C_{\varepsilon 3}$ gene. Long terminal repeat-like sequences are present in both the 5′ and the 3′ flanking regions with direct repeats. The human $C_{\varepsilon 3}$ processed pseudogene might have been created by reverse transcription of an aberrantly transcribed $C_{\varepsilon 1}$ sequence, and is located on chromosome 9 (Battey *et al.*, 1982), while the other heavy-chain constant region genes of immunoglobulin are on chromosome 14. The restriction maps of the cloned DNAs containing $C_{\varepsilon 3}$ genes of the four species are similar in the region surrounding the $C_{\varepsilon 3}$ genes as well as in the surviving coding region. These results suggest that the $C_{\varepsilon 3}$ genes of the four species evolved from a common ancestral $C_{\varepsilon 3}$ gene.

(b) *Comparison of the $C_{\varepsilon 3}$ gene sequences of chimpanzee, gorilla, orangutan and man*

The nucleotide sequence of the human $C_{\varepsilon 3}$ gene has been reported by two groups (Battey *et al.*, 1982; Ueda *et al.*, 1982). There are, however, many

† Abbreviations used: kb, 10^3 bases or base-pairs; bp, base-pairs.

Figure 2. Nucleotide sequences of the cloned $C_{\varepsilon 3}$ genes from man (HSA), chimpanzee (PTR), gorilla (GGO) and orangutan (PPY). Only nucleotides different from the human sequence are shown in the ape sequences. Deleted nucleotides are indicated by double hyphens. Hyphens indicate nucleotides not determined. Because of a typographical error, the nucleotides at positions 1404 and 1405 in the human sequence were incorrect in the previous report (Ueda *et al.*, 1982). The correct nucleotides are A and G at positions 1404 and 1405, respectively.

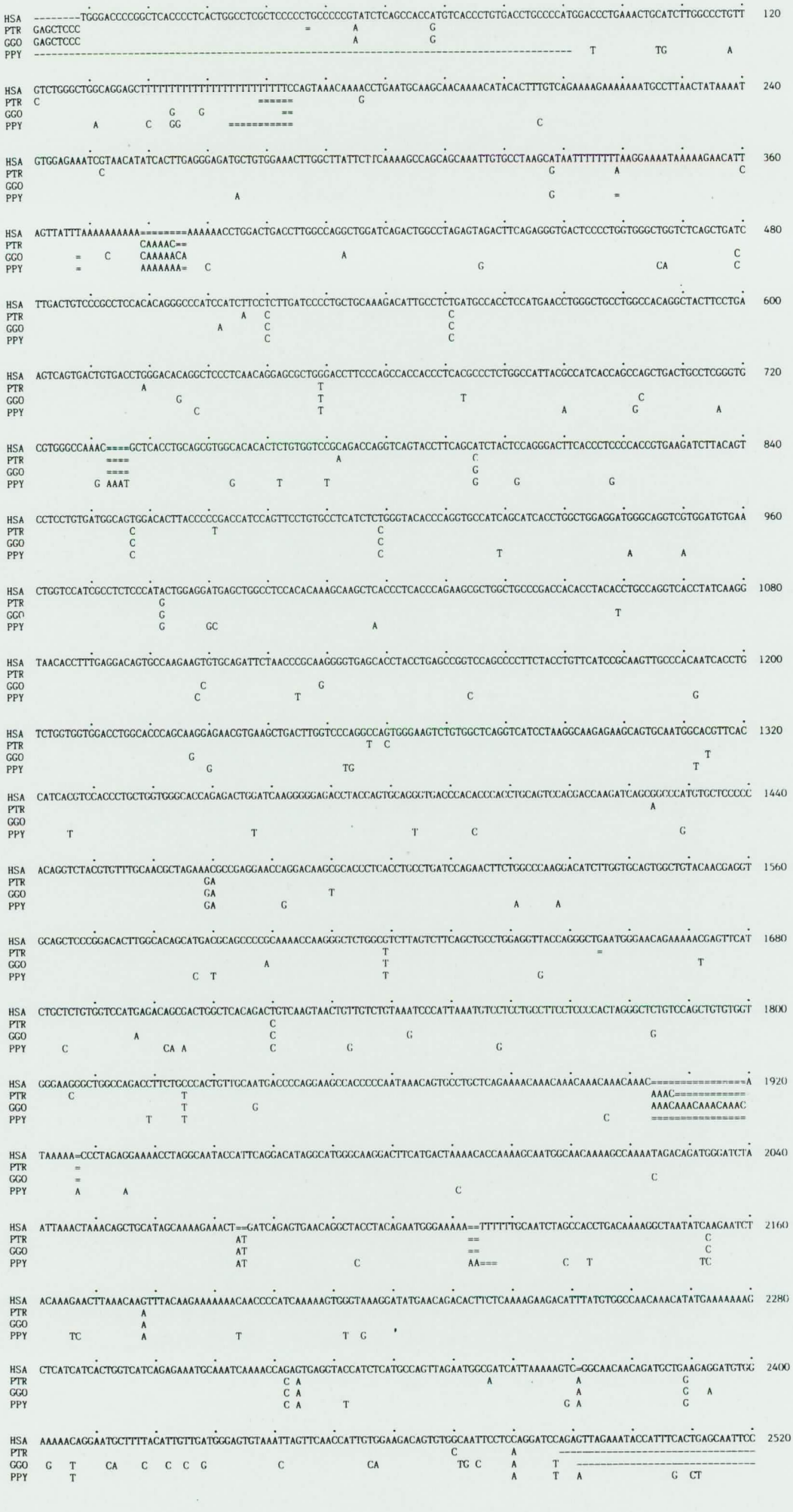

Table 1
Distance matrix among hominoids

	Human	Chimpanzee	Gorilla	Orangutan
Human		0·0139 (0·0025)	0·0175 (0·0028)	0·0406 (0·0043)
Chimpanzee	0·0139 (0·0025)		0·0162 (0·0027)	0·0383 (0·0042)
Gorilla	0·0176 (0·0028)	0·0162 (0·0027)		0·0411 (0·0044)
Orangutan	0·0408 (0·0044)	0·0384 (0·0042)	0·0413 (0·0044)	

The K^c values above and below (diagonally left to right) are those calculated by the one- and the two-parameter methods, respectively. The values in parentheses are their standard errors. In the calculation, gaps are excluded; position 92 to 2392 (2250 nucleotides) were used as described in the text.

discrepancies between the two sequences (151 positions in 1764 nucleotides commonly determined). Such a large number of discrepancies cannot be accounted for by polymorphism alone. It is unlikely that the human genome contains two separate C_ε processed pseudogenes because the restriction endonuclease maps of the flanking regions of the two C_ε genes are almost identical. The nucleotide sequence of the human $C_{\varepsilon 3}$ gene that we determined for positions 340 to 834 agreed with our previous sequence. In addition, the estimated substitution rate was unreasonably high compared with those of other pseudogenes (Koop *et al.*, 1986), when we compared the nucleotide sequence of the human $C_{\varepsilon 3}$ gene determined by Buttey *et al.* (1982) with those of the hominoid counterparts determined in the present study. We therefore adopted our sequence for the basis of comparison in the present study.

Figure 2 shows the aligned nucleotide sequences of the $C_{\varepsilon 3}$ genes of the four species. Alignment was performed so as to minimize the number of substitutions and insertions/deletions. It revealed that the gorilla sequence has an unusually high number of substitutions in the flanking region downstream from approximately position 2400. It was particularly noticeable that the nucleotide substitutions in this region of the gorilla $C_{\varepsilon 3}$ gene were much greater than those of the orangutan $C_{\varepsilon 3}$ gene in comparison with the human $C_{\varepsilon 3}$ gene, although recent data indicate that orangutan is clearly more distant from man than is gorilla. When a heterogeneity test of upstream and downstream regions by 2×2 contingency table was performed for the gorilla $C_{\varepsilon 3}$ gene, the difference in this region was extremely significant. The maximum chi-square value was obtained at position 2392, thus dividing the nucleotide sequence into two regions, one from position 92 to 2392 and the other from position 2393 to 2487. The goodness-of-fit of this division was confirmed by a homogeneity test for geometric distributions: $0{\cdot}25 < P < 0{\cdot}5$ and $0{\cdot}5 < P < 0{\cdot}75$ was obtained in position 92 to 2392 and 2393 to 2487, respectively, while $0{\cdot}01 < P < 0{\cdot}025$ was determined for the total region.

(c) *The degrees of nucleotide divergence of the $C_{\varepsilon 3}$ pseudogenes and the phylogenetic relationships among hominoids*

To construct a phylogenetic tree for the four species, the degrees of nucleotide difference between the $C_{\varepsilon 3}$ pseudogenes were calculated by one- and two-parameter methods, assuming all regions to be non-coding (Jukes & Cantor, 1969; Kimura, 1980; Kimura & Ohta, 1972). Table 1 shows the corrected values of nucleotide differences (K^c) with standard errors. The number of transitions was approximately twice that of transversions, but the K^c values calculated by the two methods were almost the same. For the following analyses we used the K^c values obtained by the one-parameter method. The K^c value for man *versus* chimpanzee was the smallest of the six combinations. On the other hand, K^c for man *versus* gorilla was nearly equal to

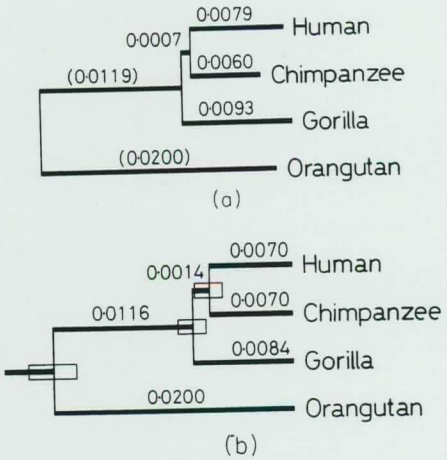

Figure 3. Phylogenetic trees among the four hominoid species constructed by the neighbour-joining (a) and the unweighted pair-group (b) methods. The values are numbers of nucleotide substitution per site for each branch. The values in parentheses are obtained by assuming the constancy of evolutionary rate between the orangutan lineage and the lineage leading to the other 3 species. The boxes represent the magnitudes of standard errors of branching points.

that for chimpanzee *versus* gorilla. K^c values for orangutan *versus* the other species were clearly distinct from those of man *versus* the African apes. This result is consistent with the generally accepted phylogenetic tree where orangutan diverged first of these four species.

Of 122 variant positions (excluding gaps), 91 were common between man and chimpanzee, 83 between man and gorilla, 33 between man and orangutan, 86 between chimpanzee and gorilla, 38 between chimpanzee and orangutan, and 32 between gorilla and orangutan. Furthermore, at three positions the four species divided into two pairs, each pair having an identical nucleotide; human and chimpanzee were paired at positions 144 and 479, and human and gorilla were paired at position 328. Both the maximum parsimony method and the compatibility method favoured the man–chimpanzee clustering over the man–gorilla clustering.

Figure 3(a) shows a phylogenetic tree drawn from the K^c values using the NJ (neighbour-joining) method (Saitou & Nei, 1987). The calculation indicates that, of the three species, chimpanzee is the closest to man. Because the NJ method produces an unrooted tree, a tentative root was given by positioning orangutan as the out-group to the other three species. The rates of nucleotide substitution in the evolutionary lineages of man, chimpanzee and gorilla seem to be more or less the same as that indicated in Figure 3(a). Note that the NJ method does not postulate the constancy of evolutionary rate. This was confirmed by the relative rate test (Wu & Li, 1985); there were no statistically significant differences in substitution rate between the three species (data not shown).

Assuming that the evolutionary rate is constant between the four species, we drew a phylogenetic tree by the unweighted pair-group method (Sokal & Sneath, 1963) as shown in Figure 3(b). Because the smallest K^c value was obtained between man and chimpanzee, these two species were clustered first. Boxes in this Figure indicate standard errors at the three branching points (Nei *et al.*, 1985). The two branching points between man and the African apes were so close that their standard errors were overlapped.

All these results taken together suggest that man and chimpanzee are more closely related to each other than either is to gorilla, in agreement with our recent qualitative study (Ueda *et al.*, 1988). During preparation of this manuscript, Miyamoto *et al.* (1987) reported a similar conclusion when comparing nucleotide sequences of $\psi\,\eta$-globin genes of man, chimpanzee, gorilla and orangutan.

Assuming that the separation of orangutan lineage occurred 14 million years ago (Pickford, 1985; Raza *et al.*, 1983; Thomas, 1985; Wu *et al.*, 1983), the divergence dates of chimpanzee and gorilla were $4{\cdot}9(\pm 0{\cdot}9)$ and $5{\cdot}9(\pm 0{\cdot}9)$ million years ago, respectively. It has been indicated that the evolutionary rate of primate genes is considerably slower than those of other mammalian genes (Ueda *et al.*, 1988; Wu & Li, 1985). Recent studies based on the nucleotide sequences of globin genes suggest that the molecular clock runs more slowly in hominoids than in other primates (Britten, 1986; Koop *et al.*, 1986). Although Li & Tanimura (1987) proposed a further slowdown of evolutionary rate in the human lineage, on the basis of Koop *et al.*'s (1986) data, the present results show a smaller number of nucleotide substitutions in the chimpanzee lineage than in the human lineage. The problem of rate constancy in the human lineage remains to be resolved because the nucleotide sequences of only a few genes have been determined in chimpanzee and gorilla, and there are no comparable data in gibbon. We need more nucleotide sequence data to test the rate constancy and to establish a more accurate local molecular clock in the hominoid lineage.

We are grateful to Dr O. Takenaka for his continuous encouragement. This investigation was supported in part by grants from the Ministry of Education, Science and Culture of Japan.

References

Battey, J., Max, E. E., McBride, W. O., Swan, D. & Leder, P. (1982). *Proc. Nat. Acad. Sci., U.S.A.* **79**, 5958–5960.

Britten, R. J. (1986). *Science*, **231**, 1393–1398.

Jukes, T. H. & Cantor, C. R. (1969). In *Mammalian Protein Metabolism* (Munro, H. N., ed.), vol. III, pp. 21–132, Academic Press, New York.

Kimura, M. (1980). *J. Mol. Evol.* **16**, 111–120.

Kimura, M. & Ohta, T. (1972). *J. Mol. Evol.* **2**, 87–90.

Koop, B. F., Goodman, M., Xu, P., Chan, K. & Slightom, J. L. (1986). *Nature (London)*, **319**, 234–238.

Li, W.-H. & Tanimura, M. (1987). *Nature (London)*, **326**, 93–96.

Messing, J. (1983). *Methods Enzymol.* **101**, 20–78.

Miyamoto, M. M., Slightom, J. L. & Goodman, M. (1987). *Science*, **238**, 369–373.

Nei, M., Stephens, J. C. & Saitou, N. (1985). *Mol. Biol. Evol.* **2**, 66–85.

Nishida, Y., Miki, T., Hisajima, H. & Honjo, T. (1982). *Proc. Nat. Acad. Sci., U.S.A.* **79**, 3833–3837.

Norrander, J., Kempe, T. & Messing, J. (1983). *Gene*, **26**, 101–106.

Pickford, M. (1985). *J. Hum. Evol.* **14**, 113–143.

Raza, S. M., Barry, J. C., Pilbeam, D., Rose, M. D., Ibrahim Shah, S. M. & Ward, S. (1983). *Nature (London)*, **306**, 52–54.

Rimm, D. L., Horness, D., Kucera, J. & Blattner, F. R. (1980). *Gene*, **12**, 301–309.

Saitou, N. & Nei, M. (1987). *Mol. Biol. Evol.* **4**, 406–425.

Sakoyama, Y., Hong, K.-J., Byun, S. M., Hisajima, H., Ueda, S., Yaoita, Y., Hayashida, H., Miyata, M. & Honjo, T. (1987). *Proc. Nat. Acad. Sci., U.S.A.* **84**, 1080–1084.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 5463–5467.

Sibley, C. G. & Ahlquist, J. E. (1987). *J. Mol. Evol.* **26**, 99–121.

Sokel, R. R. & Sneath, P. H. A. (1983). *Principles of Numerical Taxonomy*, W. H. Freeman, San Francisco.

Thomas, H. (1985). In *Ancestors* (Delson, E., ed.), pp. 42–50, Alan R. Liss Inc., New York.

Ueda, S., Nakai, S., Nishida, Y., Hisajima, H. & Honjo, T. (1982). *EMBO J.* **1**, 1539–1544.

Ueda, S., Takenaka, O. & Honjo, T. (1985). *Proc. Nat. Acad. Sci., U.S.A.* **82**, 3712–3715.

Ueda, S., Watanabe, Y., Hayashida, H., Miyata, T., Matsuda, F. & Honjo, T. (1986). *Cold Spring Harbor Symp. Quant. Biol.* **51**, 429–432.

Ueda, S., Matsuda, F. & Honjo, T. (1988). *J. Mol. Evol.* **27**, 77–83.

Wu, C.-I. & Li, W.-H. (1985). *Proc. Nat. Acad. Sci., U.S.A.* **82**, 1741–1745.

Wu, R., Xu, Q. & Lu, Q. (1983). *Acta Anthropolog. Sinica* **2**(1), 1–10 (Translated into English by Etler, D. A. (1984) *Yearbook of Phys. Anthropol.* **27**, 1–55).

Edited by B. Mach