

Nucleotide Sequence of the Beta-Globin Genes in Gorilla and Macaque: The Origin of Nucleotide Polymorphisms in Human

P. Savatier, G. Trabuchet, Y. Chebloune, C. Faure, G. Verdier, and V.M. Nigon

Département de Biologie Générale et Appliquée—UA92, Université Claude Bernard—Lyon I, 43 Boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex, France

Summary. Part of the beta-globin genes of *Macaca cynomolgus* and *Gorilla gorilla* has been cloned and sequenced. Ten putatively neutral nucleotide polymorphisms have been described at the beta-globin locus in humans. They are associated in seven combinations, which define seven different haplotypes of the beta-globin gene: four major frameworks—1, 2, 3, and 3*—and three minor frameworks, which we term KI1, KA1, and OR1. The nucleotide sequences of these frameworks are compared with those of homologous sequences in chimpanzee, colobus, macaque, and gorilla. This comparison provides strong evidence that framework 2 was the earliest framework in the human lineage. From framework 2, a rooted parsimonious tree for the six other frameworks is constructed. This phylogenetic tree is discussed in terms of the evolution of nucleotide polymorphisms as well as in terms of genetic affinities between human populations.

For each position at which there is base difference in comparing human, gorilla, and chimpanzee beta-globin genes, the phyletic lineage where the corresponding substitution occurred has been identified using the maximum parsimony procedure. The data provide evidence that polymorphisms may represent a significant component of differences between closely related species. If so, nucleotide polymorphisms may strongly bias estimates of small evolutionary distances.

Key words: DNA sequencing — DNA evolution — Macaque — Gorilla — Beta-globin gene — Nucleotide polymorphisms — Intraspecific versus interspecific variations — Hominine phylogeny

Introduction

Many studies based on DNA and protein comparisons have attempted to resolve the dichotomous branching order among the Hominoidea (human, chimpanzee, gorilla, orangutan, and gibbon). Human, chimpanzee, and gorilla are undoubtedly the three most closely related species. However, the resolving power of most techniques does not allow separation of their respective branching nodes, leading to the suspicion that the branchings occurred within a short time of one another (Sibley and Ahlquist 1984; O'Brien et al. 1985; Templeton 1985; Koop et al. 1986; Savatier et al. 1987). Mitochondrial restriction site data favor a closer relationship between chimpanzee and gorilla (Templeton 1985), but alternative phylogenies cannot be rejected. Recently, the identification of an additional immunoglobulin C-epsilon pseudogene in human and gorilla DNAs suggested that these two species are the most closely related (Ueda et al. 1985).

The first complete nucleotide sequence of the human beta-globin gene was determined by Lawn et al. (1980). Part or all of about 40 additional nucleotide sequences have since been reported from the DNA of geographically scattered individuals.

These sequences show 10 silent, putatively neutral nucleotide polymorphisms (for a review, see Orkin and Kazazian 1984; Antonarakis et al. 1985) that define four major haplotypes of the human beta-globin gene: frameworks 1, 2, 3, and 3*. Since the four frameworks are found together in different human populations, their original appearance may have predated racial divergence.

In this report, we compare the nucleotide sequences of these various human beta-globin frameworks with the homologous sequences in chimpanzee (*Pan troglodytes*, Savatier et al. 1985) and colobus monkey (*Colobus polykomos*, Martin et al. 1983). We also compare them with the homologous sequences in macaque (*Macaca cynomolgus*) and lowland gorilla (*Gorilla gorilla*), which are reported in this article. Using the maximum parsimony procedure, we constructed a nucleotide sequence for the beta-globin gene of the common ancestor of *Homo*, *Pan*, and *Gorilla*. Based on this ancestral sequence, we analyze the origin of nucleotide differences among chimpanzee, gorilla, and human at the beta-globin locus and we propose a rooted phylogenetic tree for the human beta-globin frameworks. We also analyze the effects of intraspecific nucleotide variation on estimates of evolutionary distances between closely related species. Base substitution rates and hominoid phylogeny are discussed in light of these results.

Materials and Methods

The strategy for sequencing the gorilla beta-globin gene is presented in Fig. 1. Savatier et al. (1987) gives the DNA isolation, cloning, and sequencing procedures.

Results

Nucleotide sequences of a macaque beta-globin gene and a gorilla beta-globin gene were determined and then compared with the orthologous sequences from human (Poncz et al. 1983), chimpanzee (Savatier et al. 1985), and colobus monkey (Martin et al. 1983) (Fig. 2). The comparison was restricted to a 1396-nucleotide-long segment (positions +1 to +1396 with respect to the cap site).

In this comparison, we have included putatively neutral nucleotide polymorphisms that have been identified at the beta-globin locus in human, and excluded mutations responsible for beta-chain hemoglobinopathies and beta-thalassemias. We have also excluded structural variations of the beta-globin polypeptide, even though some of them are putatively neutral (Bunn et al. 1977), because for most of them the replacement substitution has not been

identified at the DNA level. We have considered therefore only 10 neutral nucleotide polymorphisms, which have been identified from nucleotide sequencings of about 40 human beta-globin genes carried out by Lawn et al. (1980), Orkin et al. (1982), Kimura et al. (1983), Poncz et al. (1983), Kazazian et al. (1984), Orkin and Kazazian (1984), and Antonarakis et al. (1985).

1. Positions Involved in Nucleotide Polymorphisms in Humans

The 10 nucleotide polymorphisms mentioned above are shown in Fig. 2. They occur at nine nucleotide positions, of which eight display two different nucleotides and one (position +569) displays three (Figs. 2 and 3B).

Seven combinations of these polymorphisms have been identified in human populations, which define seven haplotypes or frameworks of the human beta-globin gene (Fig. 3B). Four of them (major frameworks 1, 2, 3, and 3*) are found widely among different human populations. The other three occur only once each in the 40 reported human beta-globin gene sequences. One of these three frameworks has been identified by restriction endonuclease analysis in two other individuals. We term the three minor frameworks OR1, KI1, and KA1 in reference to Orkin et al. (1982), Kimura et al. (1983), and Kazazian et al. (1984), who first identified the corresponding nucleotide polymorphisms.

The seven frameworks can be derived from each other by more than one branching route. The route illustrated in Fig. 4 corresponds to the maximum parsimony solution. Framework KI1 can be derived from either framework 1 or framework 2; the two solutions are equally parsimonious (three substitutions are necessary in both cases). The maximum parsimony network costs 10 substitutions.

Making use of the comparisons between human and nonhuman primate sequences shown in Fig. 2, we can identify framework 2 as the most primitive. At every position involved in a nucleotide polymorphism (except +59 and +511), the nucleotides found in human framework 2 are identical to the homologous nucleotides in the macaque, gorilla, chimpanzee, and colobus DNAs (Fig. 3C).

Identification of framework 2 as the most primitive framework allows the determination of the nucleotide of the ancestral human-chimpanzee-gorilla gene (which we call "beta AN") at each of the seven polymorphic positions (Fig. 3C). This ancestral sequence remains ambiguous at position +59 (T in macaque and colobus, C in gorilla and chimpanzee). At position +511, the nucleotide found in gorilla (T) is the only one different from that of framework

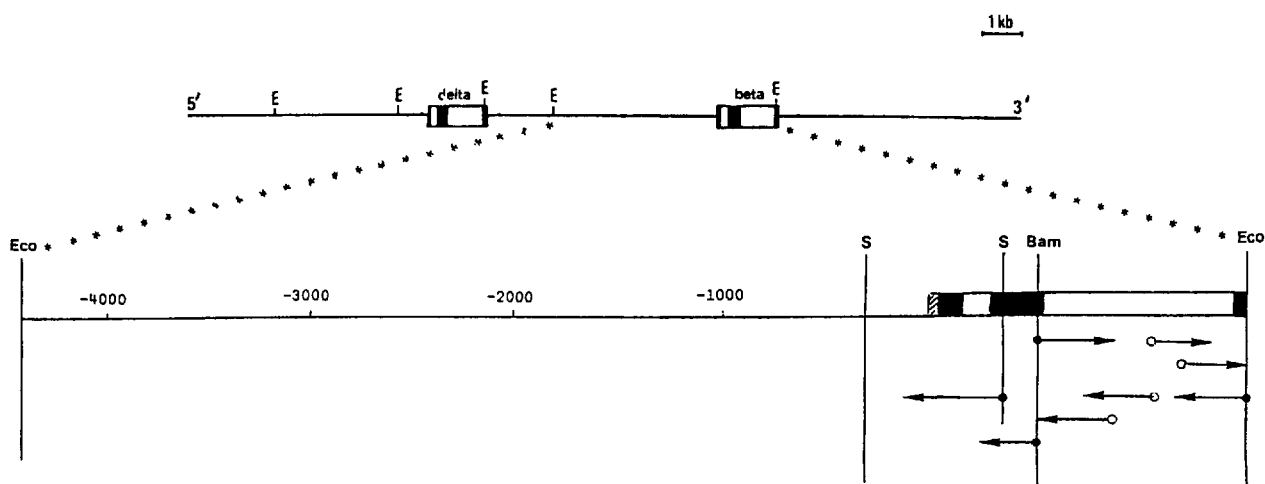


Fig. 1. Linkage maps of the gorilla delta- and beta-globin gene region deduced from blot hybridization of genomic DNA (data not shown). The map of the "Go beta EcoRI 6.4kb" clone is shown expanded. Both sequencing and restriction endonuclease mapping provided evidence that the extra 900 nucleotides found in the gorilla 6.4-kb insert with respect to both the human and chimpanzee 5.5-kb inserts come from an additional segment about 200 nucleotides 3' of the intergenic EcoRI site (data not shown). The strategy for sequencing the gorilla beta-globin gene is indicated by the horizontal arrows below the maps, indicating sequencing either from restriction endonuclease sites (filled circles) or from Bal31-generated ends (open circles). E or Eco, EcoRI; Bam, BamHI. In bar at right, the hatched area represents the 5' untranslated region, the filled areas represent introns, and the open areas represent exons

2 (C). It is also different from the nucleotide found in the other frameworks (G). Therefore, we suggest that most probably (1) C was present at position +511 in beta AN, (2) a C → T substitution occurred at this position in the gorilla lineage, and (3) a C → G substitution occurred in some humans.

A rooted parsimonious tree was constructed from the seven human beta-globin frameworks (Fig. 5). Framework 1 and framework KA1 were derived directly from the root node (framework 2) through only one substitution each. Frameworks 3*, 3, and OR1 derive from the root by three, four, and five substitutions, respectively. This implies the existence of two hypothetical intermediate frameworks between framework 2 and framework 3* that have not yet been identified in human populations. We call these frameworks H1 and H2. One of them could be easily investigated since positions +59 and +511, which define it, can be detected by the restriction endonucleases HgiA1 and AvaII, respectively.

2. Positions Not Involved in Nucleotide Polymorphisms in Humans

At 21 nucleotide positions there are base differences between framework 2 and chimpanzee or framework 2 and gorilla (Fig. 2). This implies that the 21 corresponding substitutions arose either in the phyletic lineages leading to chimpanzee or gorilla or in the lineage leading to human prior to the emergence of human beta-globin gene alleles. Most of these 21 substitution sites are in introns; five are at silent positions in coding sequences. One difference leads

to an amino acid replacement (residue position 104, Arg → Lys) between human and gorilla, in agreement with the corresponding amino acid sequences (Goodman et al. 1983).

For each of these 21 nucleotide positions, let us consider the homologous nucleotides in human, chimpanzee, macaque, gorilla, and (as far as known) in colobus (Table 1). Eighteen belong to the one-solution class. That is, for each of these 18 positions, four of the five species compared bear the same nucleotide; only one of the three species human, chimpanzee, and gorilla bears a nucleotide that is different from the one found in the other four. The maximum parsimony procedure depicts the nucleotide that is found four times (or three times when the homologous nucleotide in colobus is unknown) as already present in the last common ancestor of human, chimpanzee, and gorilla. A substitution is taken to have occurred in the phyletic lineage corresponding to the nucleotide represented only once. In the assignment of the 18 substitutions to the three phyletic lineages (Table 1), we find 4 substitutions in the gorilla lineage, 6 in the chimpanzee, and 8 in the human.

For each of the three remaining positions (+494, +715, and +1240, Table 1), the nucleotide of beta AN cannot be determined since each nucleotide is represented twice. These three sites appear to be cladistically informative. Unfortunately, each position results in a different parsimony cladogram. One (position +494) groups human with chimpanzee, another (position +715) chimpanzee with gorilla, and the third (position +1240) human with gorilla. Thus the data do nothing to resolve the trichotomy.

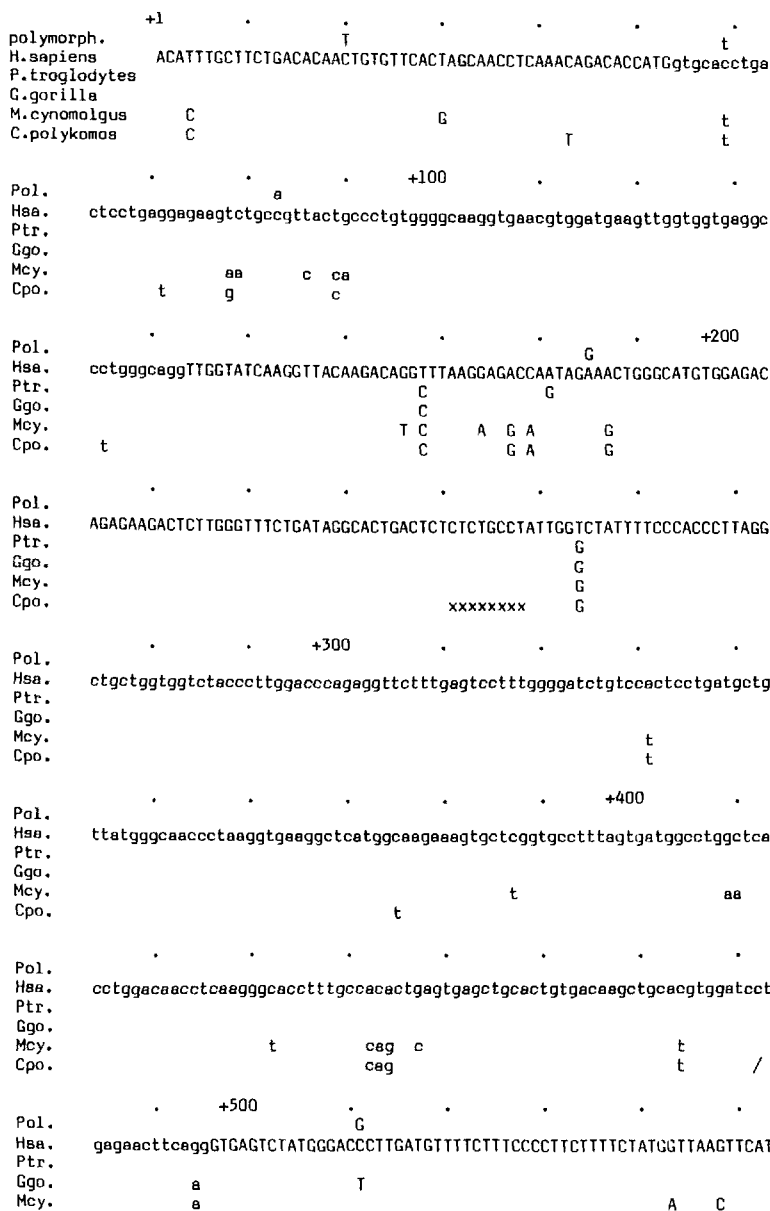


Fig. 2. Nucleotide sequence of a human (Hsa.) beta-globin gene (Poncz et al. 1983), aligned with beta-globin gene sequences from a chimpanzee (Ptr.), a gorilla (Ggo.), a colobus (Cpo.), and a macaque (Mcy.) (Martin et al. 1983; Savatier et al. 1985). Nucleotide polymorphisms (Pol.) identified in human DNA by Lawn et al. (1980), Orkin et al. (1982), Kimura et al. (1983), Poncz et al. (1983), Kazazian et al. (1984), Orkin and Kazazian (1984), and Antonarakis et al. (1985) are also shown. Numbering is from the human cap site. Gaps used to maximize identities in introns are indicated by x's, and the end of the sequenced colobus segment by a slash. Exons are reported in upper-case letters. At position +1156, the chimpanzee sequence has G instead of the A as previously reported in Savatier et al. (1985). The complete nucleotide sequence for the human beta-globin gene given by Poncz et al. (1983) is given on the second line. This nucleotide sequence is that of framework 2 (see section 1 of Results). Only nucleotide differences found between either species or alleles are indicated

Number of Nucleotide Differences Between the Ancestral "Beta AN" Gene and Each Present-Day Beta-Globin Gene

The results reported in sections 1 and 2 are summarized in Fig. 6. We shall assume the trichotomy among *Homo*, *Pan*, and *Gorilla*. In the phyletic lineage of *Homo*, we have estimated eight differences between beta AN and the stem gene (excluding the three cladistically informative positions; see Table 1). Similarly, we have estimated six nucleotide differences between the ancestral beta-globin gene and the chimpanzee beta-globin gene and five nucleotide differences between the ancestral beta-globin gene and the gorilla beta-globin gene. Finally, between the stem gene and the human beta-globin gene we have counted zero, one, three, four, or five differences, depending on the framework.

Discussion

1. Implications of Nucleotide Polymorphisms for the Estimation of Genetic Distances Between Closely Related Species

Our data give new information concerning the evolutionary distances among human, chimpanzee, and gorilla.

We find that 14-19 differences have accumulated between human and chimpanzee at the beta-globin locus depending on which human sequence is considered. This leads to base substitution frequencies ranging from 1% to 1.4%. The range would probably increase if nucleotide polymorphisms in chimpanzee DNA were known. Therefore, our data provide direct evidence that polymorphisms represent a no-

Pol. +600 . . .
 Hsa. GTCATAGGAAGGGAGAAGTAACAGGGTACAGTTAGAAATGGGAAACAGACGAATGATTGCATCAGTGTG
 Ptr. I
 Ggo. T
 Mcy. A T T G G A A

Pol.
 Hsa. GAAGTCTCAGGATCGT^{A/T}TTTAGT^TTTCTTTTATTGCTGTTCAACAATGTTTCTTTGTxxTTAATTC
 Ptr. C A
 Ggo. C A
 Mcy. C Ax G T GT

Pol. +700
 Hsa. TTGCTTTCTTTTTTTCTTCTCCGCAATTTTACTATTATACTTAATGCCTTAACATTGTGTATAACAA
 Ptr. T G
 Ggo. T C G
 Mcy. C x T T TGC A

Pol. +800
 Hsa. AAGGAAATATCTCTGAGATACATTAAGTAACTTAAAAAAXACTTTACACAGTCTGCCTAGTACATTAC
 Ptr. TG G
 Ggo. A G
 Mcy. T C GC G C

Pol. +900
 Hsa. TATTGGAATATATGTGTGCTTATTGTCATATTCATAATCTCCCTACTTTATTTCTTTTATTTTJAATT
 Ptr. A
 Ggo. A
 Mcy. G A Cx

Pol.
 Hsa. GATACATAATCATTATACATATTTATGGGTAAAGTGAATGTTTAAATATGTGTACAxxxxCATATG
 Ptr.
 Ggo.
 Mcy. A T CA A CATTG

Pol. +1000
 Hsa. ACCAAATCAGGGTAATTTTGCATTTGTAATTTTAAAAATGCTTTCTTTTAAATATACTTTTTTGTIT
 Ptr.
 Ggo.
 Mcy. T TT x GC

Pol. +1100
 Hsa. ATCTTATTTCTAATACTTTCCCTAATCTTTCTTTTCAGGGCAATAATGATACAATGTATCATGCCTCTT
 Ptr. C
 Ggo.
 Mcy. T C A CC T

Pol. C
 Hsa. TGCACCATCTAAAGAATAACAGTGATAATTTCTGGGTAAAGCAATAGCAATATTTCTGCATATAxxxx
 Ptr. G
 Ggo.
 Mcy. T AGTA

Pol. +1200
 Hsa. xAATATTTCTGCATAFAAATGTAAGTGTAAAGAGGTTTCATATTGCTAATAGCAGCTACAATCCAGC
 Ptr. T
 Ggo. T
 Mcy. G T T

Pol. +1300
 Hsa. TACCATTCTGCTTTTATTTTATGTTGGGATAAGGCTGGATTATTCTGAGTCCAAGCTAGGCCCTTTTGC
 Ptr. T
 Ggo.
 Mcy. G T T A

Pol. B
 Hsa. TAATCATGTTACACTCTTATCTTCTCCACAGctcctgggcaacgtggtgtgtgtgtggtggccca
 Ptr. G
 Ggo. G t
 Mcy. T G C g

Pol. +1396
 Hsa. tcactttggcaaga
 Ptr.
 Ggo.
 Mcy.

Fig. 2. Continued

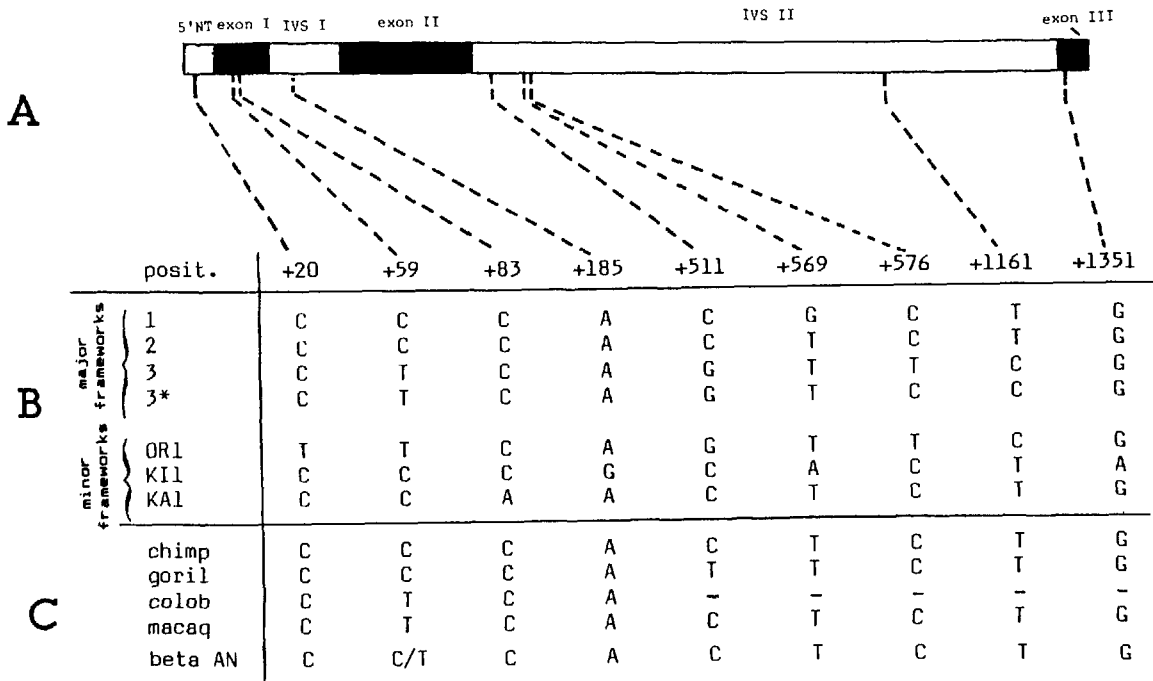


Fig. 3A-C. A Positions of the 10 nucleotide polymorphisms identified at the human beta-globin gene locus between positions +1 and +1396 (numbering is that used in Fig. 2). B Combinations of these nucleotide polymorphisms that define the seven frameworks identified in human populations. C Homologous nucleotides in the chimpanzee, gorilla, macaque, and colobus DNAs analyzed in this report. "Beta AN" is the ancestral human-gorilla-chimpanzee beta-globin gene. IVS, intervening sequence; NT, untranslated region

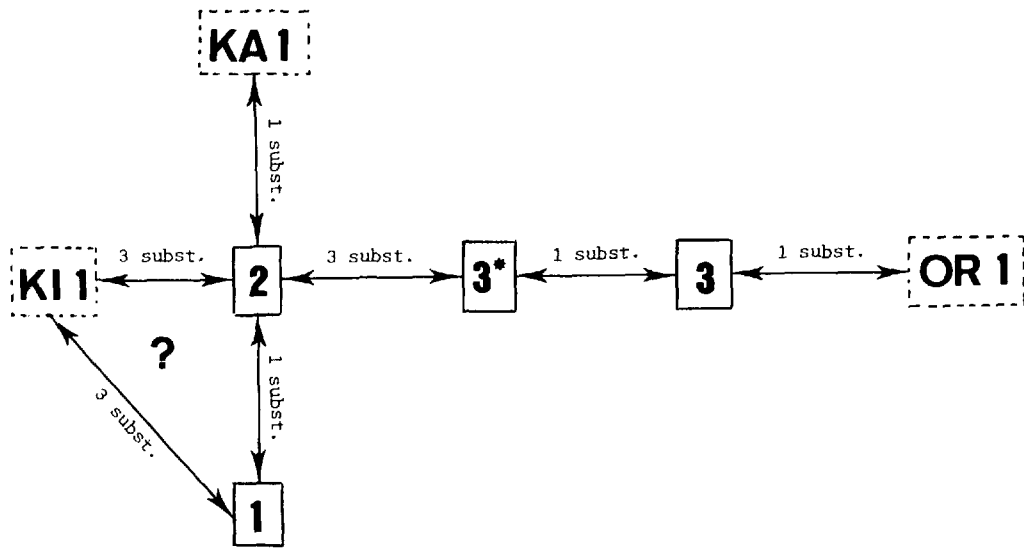


Fig. 4. Maximum parsimony network for the seven human beta-globin frameworks. Major frameworks are enclosed within a continuous line, minor frameworks within a dashed line

ticeable part of interspecific differences between closely related species; nucleotide polymorphisms thus introduce a bias into calculations of evolutionary distances between closely related homologous sequences.

In sequence comparisons among human, chimpanzee, and gorilla eta-globin pseudogenes, Chang and Slightom (1984) found that both the chimpanzee and the gorilla pseudogenes have accumulated

twice as many substitutions as their human counterpart since the time these three species began to diverge. This led these authors to conclude that the base substitution rate was lower in *Homo* than in *Pan* and *Gorilla*. In contrast, in comparing the delta-beta-globin intergenic sequences of human, chimpanzee, and macaque (Savatier et al. 1987), we did not find any difference between the base substitution frequencies in *Homo*, 0.73 ± 0.14 , and in *Pan*, 0.83

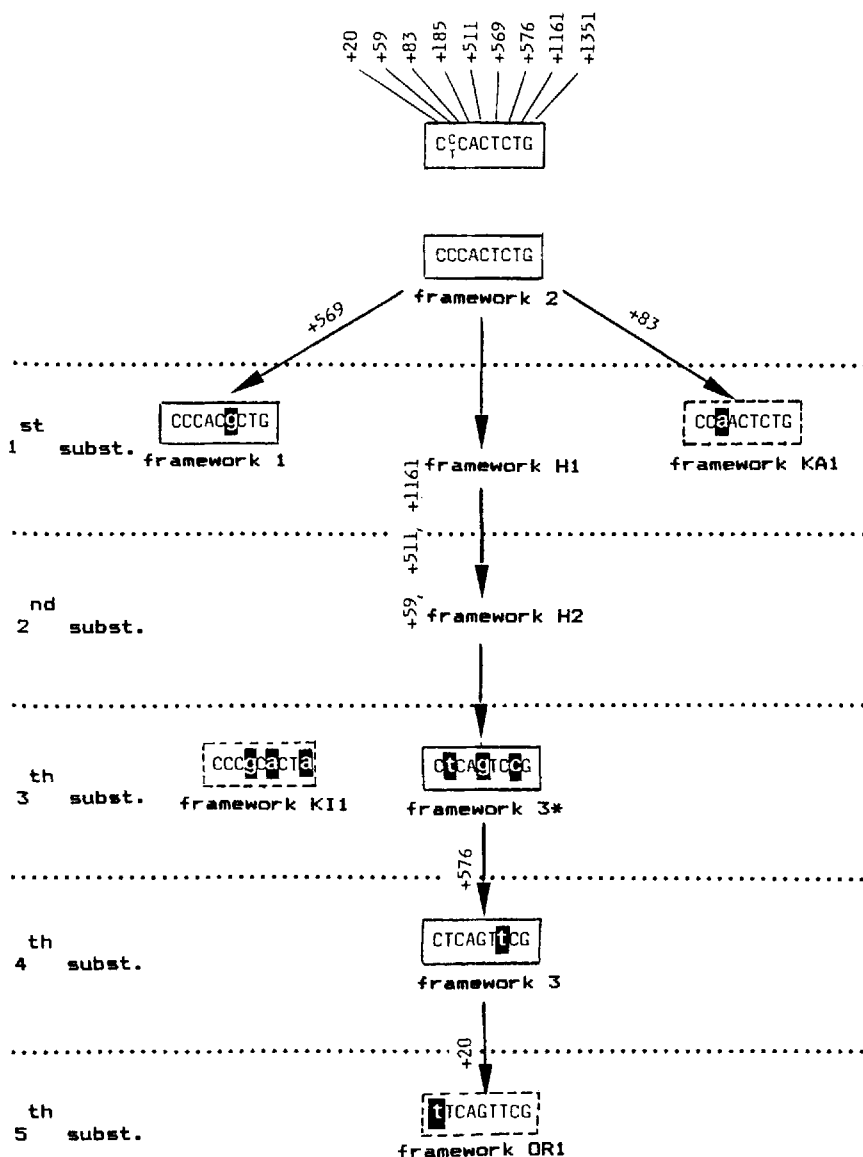


Fig. 5. Rooted parsimonious tree constructed from the seven human beta-globin frameworks. Only the nine nucleotides that define these frameworks are shown. Each connecting line between two nodes corresponds to one substitution. For each node, the position of the substitution is indicated by a lower-case white letter. Frameworks H1 and H2 are hypothetical intermediate frameworks (see text). Each framework has been classified as a one-, two-, three-, four-, or five-substitution type descendant according to the number of substitutions since the root

± 0.15 ($P > 35\%$). This discrepancy might be explained in part by nucleotide polymorphisms inside the compared regions, since both the chimpanzee and the human whose DNAs were analyzed by Chang and Slightom (1984) were different from ours. Therefore, whether base substitution rates have been equal in human and African great apes (gorilla and chimpanzee) is still an open question.

We suggest that the estimation of accurate evolutionary distances between closely related species depends on three conditions: (1) The main nucleotide polymorphisms inside the compared region must be identified. (2) The sequence comparison must be between species that are as closely related as possible. [The ambiguous status of positions +494, +715, and +1240 in our data may be due in part to the distance of macaque from hominoids; identification of the homologous nucleotides in the orangutan, which is much closer to the African great

apes than the macaque (Sibley and Ahlquist 1984; Koop et al. 1986) might make it possible to resolve these ambiguous positions.] (3) The comparison must be between large nucleotide segments so as to give more accurate estimates of base substitution frequencies.

It is noteworthy that the overall number of nucleotide polymorphisms might be much higher in humans than in chimpanzees because of the large differences in effective population size. If so, the base substitution frequencies in the *Pan* and *Homo* lineages may not be strictly comparable.

2. Branching Patterns Within Hominines Based on Nonquantitative Methods

Various quantitative methods have been used to resolve the branching order among human, gorilla, and chimpanzee based on data for both nuclear and

Table 1. Nucleotide positions showing base differences between the human framework 2 sequence and the chimpanzee or gorilla homologous sequence

Nucleotide position in the human sequence ^a	Nucleotide in						Lineage where substitution occurred
	Macaque	Chimp	Gorilla	Colobus	Human	Beta AN ^b	
+168	C	C	C	C	T	C	Human
+181	A	G	A	A	A	A	Chimp
+254	G	G	G	G	T	G	Human
+494	A	G	A	—	G	A/G	?
+627	C	C	C	—	G	C	Human
+638	A	A	A	—	G	A	Human
+715	C	T	T	—	C	C/T	?
+730	T	T	C	—	T	T	Gorilla
+748	G	G	G	—	A	G	Human
+789	T	T	A	—	T	T	Gorilla
+803	A	T	A	—	A	A	Chimp
+804	G	G	G	—	C	G	Human
+810	A	G	A	—	A	A	Chimp
+872	A	A	A	—	C	A	Human
+1048	A	C	A	—	A	A	Chimp
+1156	A	G	A	—	A	A	Chimp
+1207	G	G	T	—	G	G	Gorilla
+1240	T	T	C	—	C	T/C	?
+1250	G	T	G	—	G	G	Chimp
+1331	G	G	G	—	A	G	Human
+1357	C	C	T	—	C	C	Gorilla

^a Numbered as in Fig. 2

^b "Beta AN" denotes the deduced sequence of the beta-globin gene of the common ancestor of human, chimpanzee, and gorilla

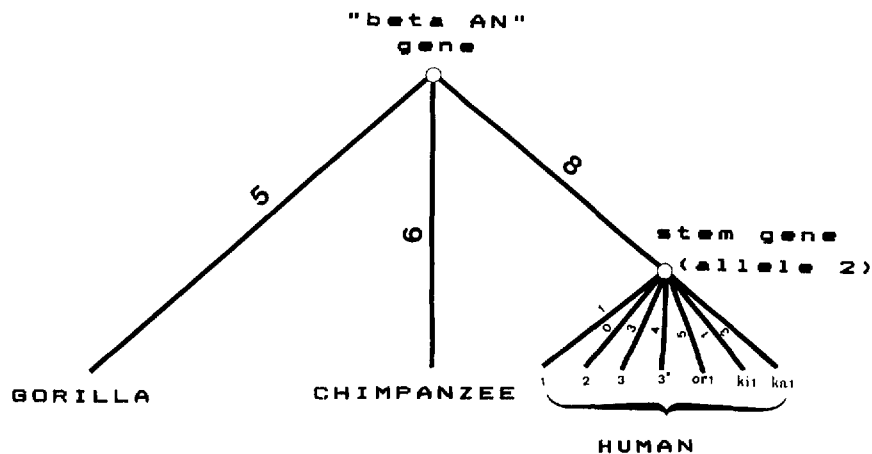


Fig. 6. Number of nucleotide differences found between each present-day beta-globin gene and either the ancestral gene beta AN or the "stem gene"

mitochondrial DNA: percentages of substitutions, restriction endonuclease mapping, thermal stability of DNA-DNA hybrids, and maximum parsimony analysis of nucleotide substitutions (for review, see Sibley and Ahlquist 1984; Templeton 1985; Koop et al. 1986; see also Savatier et al. 1987). Data were found to support each of the three alternative cladograms but most of these methods were not able to separate the divergence nodes with statistical significance.

Why is it that there is not better agreement among these different experimental approaches? First, probably no more than 2 million years separate the

human-chimpanzee, human-gorilla, and chimpanzee-gorilla divergence nodes; this represents less than 20% of the time since human-chimpanzee-gorilla divergence [about 8 million years according to Sibley and Ahlquist (1984); see also Koop et al. (1986)]. Since the accuracy of calculated evolutionary distances is limited either by the length of the compared nucleotide sequences or by the variability of the experimental results (as in DNA-DNA hybridization), the resolving power of any one of these techniques is not sufficient to separate the divergence nodes with statistical significance. Second, we have shown that nucleotide polymorphisms might rep-

resent a noticeable part of the interspecific differences among human, gorilla, and chimpanzee at the beta-globin locus. Templeton (1985) has previously noted that these interspecific differences are not evolutionarily significant unless they are much larger than the intraspecific differences.

In this context, nonquantitative methods have been interpreted as more reliable. Ueda et al. (1985) found that a truncated immunoglobulin epsilon pseudogene (C epsilon 2) is present in human and gorilla but absent from the DNAs of all other catarrhines analyzed in their report. This led them to conclude that gorilla is more closely related to human than is chimpanzee.

In the nucleotide sequence of the beta-globin gene, we found three cladistically informative positions. Unfortunately, these three positions lead to the three alternative hypotheses for the branching order among *Homo*, *Pan*, and *Gorilla*. The base difference found between primates at position +494 (A in both macaque and gorilla, G in both human and chimpanzee) leads to the single amino acid replacement (residue position 104) found between the gorilla and human and chimpanzee beta-globin polypeptides. Similarly, human and chimpanzee have arginine and aspartic acid at position A-gamma-104 and alpha-23, respectively; gorilla and all other catarrhines that have been analyzed have lysine and glutamic acid, respectively, at these positions (Goodman et al. 1983). This led Goodman et al. (1983) to conclude that human and chimpanzee probably are more closely related than are human and gorilla or chimpanzee and gorilla.

Ueda's and Goodman's conclusions can be reconciled if one assumes that the *Homo*, *Pan*, and *Gorilla* lineages arose from a common population in which the above-mentioned differences were already present as polymorphisms (i.e., both presence and absence of the C epsilon 2 immunoglobulin pseudogene, both arginine and lysine at positions beta-104 and A-gamma-104, both aspartic and glutamic acid at position alpha-23). The branching of *Homo*, *Pan*, and *Gorilla* might have preserved different combinations of these ancestral polymorphisms and discarded others. In this case, one might expect to draw different conclusions about their dichotomous branching order depending on the chromosome or recombination region analyzed.

3. Origin and Phylogeny of the Human Beta-Globin Frameworks

The data presented here provide strong evidence that framework 2 was the stem or founder framework in which successive substitutions led to the present-day polymorphisms at the beta-globin locus in humans. Other polymorphisms might have ex-

isted prior to the diversification of this stem framework. All but one (framework 2) apparently disappeared, although we cannot exclude that some may still be present at low frequencies.

Major frameworks have been identified in most human populations so far analyzed (American black, Mediterranean, Indian, Chinese, and Cambodian), which suggests that their appearance predated racial divergence. However, framework 1 is present in more than 95% of Africans (Pagnier et al. 1985). On the other hand, framework 3 has been identified only in Mediterraneans and American blacks, and framework 3* only in South Asian populations. Since framework 3* most probably gave rise to framework 3 (Fig. 5), it might be expected that the former would be as widely distributed as the latter. Two hypotheses can be considered to explain the discrepancy:

1. Genetic distances measured on the basis of both nucleotide polymorphisms at protein loci and restriction enzyme polymorphisms provide evidence that the Negroid group and Caucaso-Mongoloid group diverged from each other first and that the Caucasoid and Mongoloid groups diverged from each other later (Nei and Roychoudhury 1972, 1982; Wainscoat et al. 1986). This proposed phylogeny requires framework 3 to have arisen prior to the Negroid-Caucaso-Mongoloid split and framework 3* to have disappeared independently from both the Negroid and Caucasoid populations.

2. According to the most parsimonious solution, both the appearance of framework 3 and disappearance of framework 3* took place in a Negroid-Caucasoid stem population prior to the Negroid-Caucasoid split. This implies that the Mongoloid group and the Negroid-Caucasoid group diverged from each other first and the Caucasoid and Negroid groups from each other later. Previously, only blood groups and histocompatibility locus antigens had supported this phylogeny (Cavalli-Sforza and Edwards 1964; Piazza et al. 1975).

The minor frameworks (OR1, KI1, and KA1) have been identified only once each in about 40 beta-globin gene sequences. In the parsimonious network (Fig. 4), these minor frameworks appear on exterior branches and in no case do they constitute intermediate forms. Both observations suggest them to be of more recent origin than the major frameworks.

Unidentified minor beta-globin frameworks probably exist in human populations. The parsimonious tree presented in Fig. 5 assumes the existence of additional frameworks as intermediates. Only one of them could be easily investigated (framework H1 or H2) since both positions that define it are detectable by restriction endonuclease mapping. However, sequence data suggest that only 10% of individuals may be minor-framework car-

riers, which implies that their detection depends on the improvement of methods that screen genomic DNA directly rather than on sequencing. If intermediate frameworks H1 and H2 still exist, their geographical location might provide information on the prehistoric migrations and genetic affinities of human populations.

How old is the human stem gene? Assuming that an average of two positions have undergone nucleotide substitutions between the stem gene and the present-day frameworks and that nine substitutions have been fixed between beta AN and the stem gene, the evolutionary distance between the stem gene and the present genes would represent about 20% of the human-gorilla-chimpanzee divergence time of about 8 million years, i.e., approximately 1.5 million years. Nei and Roychoudhury (1982) estimated that the Negroid lineage separated from the Caucaso-Mongoloid lineage 120,000 years ago. Therefore, assuming that all four major frameworks arose prior to the Negroid-Caucaso-Mongoloid split, our estimate of 1.5 million years is fully consistent with Nei's.

4. The Evolution of Polymorphisms

As already mentioned, our data show the absence of intermediate forms between beta AN and the stem gene as well as between the stem gene and the present-day framework 3*. This situation might be a direct consequence of genetic drift. Considering the limited number of alleles that can be maintained in a finite population, the spreading of new alleles and the disappearance of older ones, whether by neutral drift or by linkage of a polymorphic form to a selectively advantageous mutation capable of inducing changes in allele frequencies and loss of genetic variability, will necessarily be concurrent.

Acknowledgments. We are grateful to Dr. Ivanoff (Centre de Recherches Médicales de Franceville, Gabon) for providing us blood samples from gorilla. This work was supported by research grants from INSERM, CNRS, and the Commission of the European Communities.

References

- Antonarakis SE, Kazazian HH, Orkin SH (1985) DNA polymorphism and molecular pathology of the human globin gene cluster. *Hum Genet* 69:1-14
- Bunn HF, Forget BG, Ranney HN (1977) *Human hemoglobins*. WB Saunders, Philadelphia
- Cavalli-Sforza LL, Edwards AWF (1964) *Genetics today*. Pergamon, Oxford
- Chang LYE, Slightom JL (1984) Isolation and nucleotide sequence analysis of the beta type globin pseudogene from human, gorilla and chimpanzee. *J Mol Biol* 180:767-784
- Goodman M, Braunitzer G, Stangl A, Schrank B (1983) Evidence on human origins from haemoglobins of African apes. *Nature* 303:546-548
- Kazazian HH Jr, Orkin SH, Antonarakis SE, Sexton JP, Boehm CD, Goff SC, Waber PG (1984) Molecular characterization of seven beta-thalassemia mutations in Asian Indians. *EMBO J* 3:593-596
- Kimura A, Matsunaga E, Ohta Y, Fujiyoshi T, Matsuo T, Nakamura T, Imamura T, Yanase T, Takagi Y (1983) Structure of cloned delta-globin genes from a normal subject and a patient with delta-thalassemia: sequence polymorphisms found in the delta-globin gene region of Japanese individuals. *Nucleic Acids Res* 10:5725-5732
- Koop BF, Goodman M, Xu P, Chen K, Slightom JL (1986) Primate eta-globin DNA sequences and man's place among the great apes. *Nature* 319:234-237
- Lawn RM, Efstratiadis A, O'Connell C, Maniatis T (1980) The nucleotide sequence of the human beta-globin gene. *Cell* 21:647-651
- Martin SL, Vincent KA, Wilson AC (1983) Rise and fall of the delta globin gene. *J Mol Biol* 164:513-528
- Nei M, Roychoudhury AK (1972) Gene differences between Caucasian, Negro and Japanese populations. *Science* 177:434-436
- Nei M, Roychoudhury AK (1982) Genetic relationship and evolution of human races. *Evol Biol* 14:1-59
- O'Brien SJ, Nash WG, Wildt DE, Bush ME, Benveniste RE (1985) A molecular solution to the riddle of the giant panda's phylogeny. *Nature* 317:140-144
- Orkin SH, Kazazian HH (1984) The mutation and polymorphism of the human beta-globin gene and its surrounding DNA. *Annu Rev Genet* 18:131-171
- Orkin SH, Kazazian HH, Antonarakis SE, Goff SC, Boehm CD, Sexton JP, Waber PG, Giardina JV (1982) Linkage of beta-thalassemia mutations and beta-globin gene polymorphisms with DNA polymorphisms in human beta-globin gene cluster. *Nature* 296:627-631
- Pagnier J, Mears JG, Dunda-Belkhdja O, Schaefer-Rego KE, Beldjord C, Nagel RL, Labie D (1985) Evidence for the multicentric origin of the sickle cell hemoglobin gene in Africa. *Proc Natl Acad Sci USA* 81:1771-1773
- Piazza A, Sgaranello-Zonta L, Gluckman P, Cavalli-Sforza LL (1975) The 5th histocompatibility workshop gene frequency data: phylogenetic analysis. *Tissue Antigens* 5:445-463
- Poncz M, Schwartz E, Ballantine M, Surrey S (1983) Nucleotide sequence analysis of the delta-beta globin gene region in humans. *J Biol Chem* 258:11599-11609
- Savatier P, Trabuchet G, Faure C, Chebloune Y, Gouy M, Verdier G, Nigon VM (1985) Evolution of the primate beta-globin gene region: high rate of variation in CpG dinucleotides and in short repeated sequences between man and chimpanzee. *J Mol Biol* 182:21-29
- Savatier P, Trabuchet G, Chebloune Y, Faure C, Verdier G, Nigon VM (1987) Nucleotide sequence of the delta-beta-globin intergenic segment in the macaque: structure and evolutionary rates in higher primates. *J Mol Evol* 24:297-308
- Sibley CG, Ahlquist JE (1984) The phylogeny of the hominoid primates as indicated by DNA-DNA hybridization. *J Mol Evol* 20:2-15
- Templeton AR (1985) The phylogeny of the hominoid primates: a statistical analysis of the DNA-DNA hybridization data. *Mol Biol Evol* 2:420-433
- Ueda S, Takenaka O, Honjo T (1985) A truncated immunoglobulin epsilon pseudogene is found in gorilla and man but not in chimpanzee. *Proc Natl Acad Sci USA* 82:3712-3715
- Wainscoat JS, Hill AVS, Boyce AL, Fint J, Hernandez M, Thein SL, Old JM, Lynch JR, Falusi AG, Weatherall DJ, Clegg JB (1986) Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms. *Nature* 319:491-493