

Chimpanzee Fetal $G\gamma$ and $A\gamma$ Globin Gene Nucleotide Sequences Provide Further Evidence of Gene Conversions in Hominine Evolution¹

Jerry L. Slightom,* L.-Y. Edward Chang,* Ben F. Koop,†
and Morris Goodman†

*Department of Genetics, University of Wisconsin-Madison; and †Department of Anatomy, Wayne State University School of Medicine

The fetal globin genes $G\gamma$ and $A\gamma$ from one chromosome of a chimpanzee (*Pan troglodytes*) were sequenced and found to be closely similar to the corresponding genes of man and the gorilla. These genes contain identical promoter and termination signals and have exons 1 and 2 separated by the conserved short intron 1 (122 bp) and exons 2 and 3 separated by the more rapidly evolving, larger intron 2 (893 bp and 887 bp in chimpanzee $G\gamma$ and $A\gamma$, respectively). Each intron 2 has a stretch of simple sequence DNA (TG)_n serving possibly as a "hot spot" for recombination. The two chimpanzee genes encode polypeptide chains that differ only at position 136 (glycine in $G\gamma$ and alanine in $A\gamma$) and that are identical to the corresponding human chains, which have aspartic acid at position 73 and lysine at 104 in contrast to glycine and arginine at these respective positions of the gorilla $A\gamma$ chain. Phylogenetic analysis by the parsimony method revealed four silent (synonymous) base substitutions in evolutionary descent of the chimpanzee $G\gamma$ and $A\gamma$ codons and none in the human and gorilla codons. These Hominae (*Pan*, *Homo*, *Gorilla*) coding sequences evolved at one-tenth the average mammalian rate for nonsynonymous and one-fourth that for synonymous substitutions. Three sequence regions that were affected by gene conversions between chimpanzee $G\gamma$ and $A\gamma$ loci were identified: one extended 3' of the hot spot with $G\gamma$ replaced by the $A\gamma$ sequence, another extended 5' of the hot spot with $A\gamma$ replaced by $G\gamma$, and the third conversion extended from the 5' flanking to the 5' end of intron 2, with $G\gamma$ replaced here by the $A\gamma$ sequence. A conversion similar to this third one has occurred independently in the descent of the gorilla genes. The four previously identified conversions, labeled C1-C4 (Scott et al. 1984), were substantiated with the addition of the chimpanzee genes to our analysis (C1 being shared by all three hominines and C2, C3, and C4 being found only in humans). Thus, the fetal genes from all three of these hominine species have been active in gene conversions during the descent of each species.

Introduction

The β -type hemoglobin genes of either human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), or lowland gorilla (*Gorilla gorilla gorilla*) occur as a cluster spanning a chromosomal region of ~ 50 kb, with the order of the genes and developmental expression being from 5' to 3': ϵ (embryonic), duplicated γ (fetal), the pseudogene now labeled $\psi\eta$ (Goodman et al. 1984; Harris et al. 1984), and δ and β (adult) (Fritsch et al. 1980; Barrie et al. 1981; Zimmer 1981; Chang and Slightom 1984; Scott et al.

1. Key words: chimpanzee fetal globin genes, DNA sequencing, gene evolution, gene conversion.

Current addresses: Jerry L. Slightom, The Upjohn Company, Kalamazoo, Michigan 49001; and L.-Y. Edward Chang, The Public Health Research Institute of the City of New York, Incorporated, 455 First Avenue, New York, New York 10016.

Address for correspondence and reprints: Dr. Jerry L. Slightom, The Upjohn Company, Kalamazoo, Michigan 49001.

Mol. Biol. Evol. 2(5):370-389, 1985.

© 1985 by The University of Chicago. All rights reserved.
0737-4038/85/0205-0355\$02.00

1984). The two γ loci result from a duplication that occurred ~ 35 Myr ago in a basal member of the primate infraorder Catarrhini, i.e. in a common ancestor of Cercopithecoidea (old world monkeys) and Hominoidea that had already separated from Platyrrhini (new world monkeys) (Barrie et al. 1981; Shen et al. 1981). The polypeptide products from the two γ loci in humans, chimpanzees, and gorillas can be distinguished by an amino acid replacement in position 136, the 5' locus encoding glycine and the 3' locus encoding alanine (Schroeder et al. 1963; DeJong 1971; Huisman et al. 1973; Slightom et al. 1980; Scott et al. 1984).

The complete nucleotide sequence of the human fetal globin genes has been determined by Slightom et al. (1980), who sequenced both the $G\gamma$ and $A\gamma$ genes known to be linked on the same chromosome (designated A) and the allelic $A\gamma$ gene from the opposite chromosome (designated B). Comparisons of these three sequences showed that a portion of the $A\gamma$ gene sequence more closely resembles the sequence of the $A^{G\gamma}$ gene than it does its $B\gamma$ allele. Sequence identity between these nonallelic genes covers $\sim 1,500$ bp of DNA, starting in intron 2, 600 bp 3' from exon 2, extending through exons 1 and 2 and intron 1, and ending in the 5' flanking region (Shen et al. 1981). However, the 3' third of these two $A\gamma$ genes shows the identity relationship expected of allelic genes by being more alike than the nonallelic $A^{G\gamma}$ and $A\gamma$ genes. These results suggest that sequences from the $A^{G\gamma}$ gene (donor sequence) had been superimposed onto the $A\gamma$ gene (acceptor sequence) by a mechanism involving gene conversion (Slightom et al. 1980). This gene conversion mechanism is believed to involve a stretch of simple-sequence DNA $(TG)_n$ located at the 3' boundary of this converted region, ~ 600 bp downstream (3') from the exon 2/intron 2 junction. It has been suggested that this simple-sequence DNA provides a "hot spot" for strand breakage and transfer between the duplicated γ genes (Slightom et al. 1980). The value of n for these hot spots ranges from 12 to 22 for the human γ genes.

Scott et al. (1984) sequenced the fetal globin genes from the gorilla and found them to be very similar to the human genes, including the presence of the hot spot in intron 2 ($n = 24$ for $G\gamma$ and 13 for $A\gamma$). Detailed sequence comparisons among the human γ genes (both chromosomes A and B) and gorilla $G\gamma$ and $A\gamma$ genes supported the conversion identified by Slightom et al. (1980) and suggested that four conversion events, referred to as C1–C4, also affected either human or human and gorilla genes. Conversion C1 occurred prior to separation of the human and gorilla lineages, ~ 10 Myr ago. Conversions C2 and C3 occurred in the protohumans several million years ago, sometime after separation of the gorilla and human lineages. The latest human conversion, C4, first described by Slightom et al. (1980), is thought to have occurred ~ 1 Myr ago.

The present study extends our investigation of primate fetal globin genes to another primate species closely related to man, namely, the chimpanzee (*Pan troglodytes*). Genealogically, *Pan*, *Homo*, and *Gorilla* may be classified as members of the subfamily Homininae (Goodman 1976; Goodman et al. 1983). Comparisons of the fetal globin gene sequences of these three hominines (human, chimpanzee, and gorilla) and phylogenetic analysis by the parsimony method, with the rabbit $\gamma:\beta 3$ gene sequence (Hardison 1981) serving as the nearest outgroup (see below), suggest that, after divergence of *Pan* from *Homo*, three conversion events occurred between the chimpanzee $G\gamma$ and $A\gamma$ sequences. The hot spot sequence found in intron 2 of the chimpanzee genes is located at the borders of two of the three chimpanzee conversions. The $(TG)_n$ simple sequence is a member of a highly repeated family that is distributed throughout the genome of many eukaryotic species (Miesfield et al. 1981), and such sequences may

participate in many recombinational events. However, our data also show that in the separate lineages to gorilla and chimpanzee, the 5' sides of the $^G\gamma$ and $^A\gamma$ genes may have been involved in conversion events that were not bordered by the $(TG)_n$ hot spot. Thus the presence of a $(TG)_n$ hot spot may not be necessary for conversion; however, its presence may enhance the frequency of particular gene conversions.

Material and Methods

Material

Restriction endonucleases *EcoRI*, *BamHI*, *PstI*, *AvaII*, *BglII*, *HindIII*, *Sau3A*, *SacI*, *XbaI*, and *XhoI* were obtained from Promega-Biotec. Polynucleotide kinase came from P-L Biochemicals, and bovine intestinal alkaline phosphatase from Boehringer-Mannheim. (α - ^{32}P) dCTP (800 Ci/mM) and (γ - ^{32}P) ATP (2,000–3,000 Ci/mM; $1 \mu\text{Ci} = 3.7 \times 10^{10}$ Bq) and T_4 ligase were obtained from New England Nuclear. Chemicals used for DNA sequencing were obtained from vendors recommended by Maxam and Gilbert (1980). X-ray film on rolls (20 cm \times 25 m; XAR-351) and sheets (35 \times 45 cm; XAR-5) were from Kodak. Intensifying screens (Quanta III; 35 cm \times 1 m) were from DuPont. Nitrocellulose paper (BA-85) came from Schleicher and Schuell, and 3 MM paper from Whatman.

DNA Isolation and Cloning

Total DNA from chimpanzee 1 (*Pan troglodytes*), prepared from blood samples according to the method of Kan and Dozy (1978), was a gift from Dr. E. Zimmer (Louisiana State University, Baton Rouge, La). Purified DNA was partially digested with either *Sau3A* or *EcoRI*, and fragments of 15–20 kbp were size selected on 5% 20% NaCl gradients (Slightom et al. 1980). Sized DNA fragments were cloned into the appropriate lambda vector arms, *EcoRI*-cut fragments into *EcoRI*-cut Charon 32 arms, (Loenen and Blattner 1983) and *Sau3A*-cut fragments into *BamHI*-cut Charon 30 arms (Rimm et al. 1980). Recombinant-phage DNAs were packaged into phage capsids using the in vitro phage-packaging procedure described by Hohn (1979). Charon 30 recombinant phages were plated on *E. coli* strain DP50SupF (Blattner et al. 1977), and Charon 32 phages were plated on a *RecA*⁻ *E. coli* strain ED8767 (Murray et al. 1977). The resulting recombinant-phage libraries, $1-2 \times 10^6$ phages, were screened by the method described by Benton and Davis (1977) using the ^{32}P nick-translated (Maniatis et al. 1975), 245-bp *AvaII*-*EcoRI* fragment from the γ -globin cDNA clone pJW151 (Wilson et al. 1978) as probe. Lambda clones containing chimpanzee fetal-globin-gene regions were purified, and phage growths and DNA isolations were done as described by Slightom et al. (1980). Two recombinant clones, which contain parts of the chimpanzee fetal-globin-gene duplication region, were isolated and designated Ptr Ch30-12.6 and Ptr Ch32-8.1. Clone Ptr Ch30-12.6 contains the $^A\gamma$ gene and also the β pseudogene named $\psi\eta$ (Chang and Slightom 1984; Goodman et al. 1984; Harris et al. 1984), and clone Ptr Ch32-8.1 contains the linked $^G\gamma$ and $^A\gamma$ fetal genes. The 7.0-kbp *EcoRI* and both 2.64-kbp *EcoRI* fragments were isolated and cloned into the *EcoRI* site of pBR322. Transformation of *E. coli* K12 strain HB101 was done using the extended Ca shock method described by Dagert and Ehrlick (1979). Bacterial clones containing the 7.0-kbp $^G\gamma$ or 2.64-kbp $^A\gamma$ *EcoRI* fragments are referred to as Ptr 8.1-p7.0, Ptr8.1-p2.64, and Ptr 12.6-p2.64, respectively. Plasmid DNAs were purified using the alkaline-extraction procedure described by Birnboim and Doly (1979) and following it with two gradient bandings in CsCl ethidium bromide.

DNA Sequence Analysis

Nucleotide sequences were determined using the chemical procedure described by Maxam and Gilbert (1980) and the DNA-sequencing gel techniques described by Chang and Slightom (1984). DNA sequences were assembled with computer programs Gap and Map, which were supplied by the University of Wisconsin Genetics Computer Group (Devereux et al. 1984).

Phylogenetic Tree Construction

Sixty-five gene sequences, including chimpanzee $G\gamma$ and $A\gamma$ and human myoglobin (Weller et al. 1984) and consisting of just the coding regions (exons) of expressed globin genes and the corresponding regions of globin pseudogenes (Goodman et al. 1984), were studied. For genes known to be hybrids, such as lemur $\psi\eta\delta$ and mouse $\beta h0$ and $\beta h1$ (found to be $\gamma\epsilon$ hybrids), exon 3 was treated as a separate sequence from the exon 1 + 2 sequences. Also, because recombination between the $G\gamma$ and $A\gamma$ genes of the hominine species may have resulted in sequences at the 5' and 3' sides of these genes having different evolutionary histories, exon 3 was treated as a sequence separate from the exon 1 + 2 sequence. As in our previous study (Goodman et al. 1984), corresponding exon sequences could readily be aligned against one another. Alignment of complete $G\gamma$ and $A\gamma$ gene sequences from the chimpanzee, man, and the gorilla (the total alignment requiring 1,843 nucleotide positions extending from 110 bases upstream of exon 1 to 261 bases downstream of exon 3) was also easily achieved by using gaps as needed to increase the number of matching bases. The rabbit $\gamma:\beta 3$ globin sequence (Hardison 1981) was used as the nearest outgroup to root the branching arrangement of minimum length at each nucleotide position that varied among the compared hominine sequences. The alignment of rabbit $\gamma:\beta 3$ and gorilla $A\gamma$ sequence was obtained using the algorithm of Smith and Waterman (1981), which we modified (Goodman et al. 1984).

The most parsimonious phylogenetic tree for the coding (exon) sequences was constructed as previously described (Goodman et al. 1984). The boundaries of putative gene conversion events were identified after determining the most parsimonious branching arrangement for each position contained in a file of complete coding and noncoding sequences (the rabbit and seven hominine genes). On each side of a boundary, a different branching arrangement for the sequences (i.e., different evolutionary history) is depicted by the parsimony method. A possible conversion is postulated in a region of a sequence if two conditions are met. (1) Three or more substitutions and/or insertion/deletion events are required in its descent from its immediate ancestor, and (2) at least three fewer events would be necessary if the region had instead descended from the homologous region of a paralogous gene in the same taxon. If the conversion of the same region of an orthologous gene is postulated for two sister taxa, the conversion is instead postulated to have occurred in their common ancestor. Whenever this parsimony criterion did not identify a sharp boundary for the converted region, the additional criterion of high mutational divergence between the two tandem non-allelic loci was employed to suggest possible conversion boundaries. For example, position 557 could be taken as the 3' boundary of the hypothesized conversion between the gorilla nonallelic loci because a 3-bp deletion starting at position 558 distinguishes gorilla $A\gamma$ from $G\gamma$.

Results and Discussion

Isolation of Linked Fetal Globin Genes From the Chimpanzee

From a Charon 30 recombinant-phage library of chimpanzee DNA, three clones containing γ genes were isolated. Detailed restriction-enzyme site mapping showed, however, that two of these clones had deleted the $\Lambda\gamma$ gene, probably as a result of *E. coli*-mediated recombinational events that took place during propagation of these phages (data not shown). Instability of linked fetal globin genes in lambda clones had been observed during the isolation of γ genes from the gorilla (Chang and Slightom 1984). The third Charon 30 clone, Ptr Ch30-12.6, was found to be quite stable during propagation and to contain the chimpanzee $\Lambda\gamma$ fetal gene linked to the $\psi\eta$ gene, (see fig. 1).

Deletion of the linked fetal globin genes during phage propagation was avoided by constructing a chimpanzee-DNA phage library in which partially *EcoRI*-digested fragments were cloned into the lambda vector Charon 32. Charon 32 was selected because it grows well on recombination-deficient *E. coli* hosts (Loenen and Blattner 1983). Clone Ptr Ch32-8.1, containing the linked chimpanzee $G\gamma$ and $\Lambda\gamma$ fetal globin genes, was isolated (see fig. 1) without any difficulty or indication that its insert is unstable during phage propagation.

The $G\gamma$ and $\Lambda\gamma$ fetal globin genes from clones Ptr Ch30-12.6 and Ptr Ch32-8.1 were subcloned into pBR322 and are designated Ptr 8.1-p7.0 ($G\gamma$ gene), Ptr 8.1-p2.64, and Ptr 12.6-p2.64 ($\Lambda\gamma$ gene). Restriction-enzyme mapping of these subclones shows that most of the notable sites found in the human genes are also found in these chimpanzee genes, with the exception of three sites; the *EcoRI* site 3' of the $G\gamma$ gene, two *XhoI* sites (one in intron 2 of each gene), and the polymorphic *HindIII* site (Jeffreys 1979) are absent from the clones of the fetal globin gene of the chimpanzee (see fig. 1).

Nucleotide Sequence of Chimpanzee Fetal Globin Genes

The strategy used to sequence the chimpanzee fetal globin genes is shown below the maps in figure 1. The nucleotide sequence of the $G\gamma$ gene was determined from clone Ptr Ch32-8.1, starting 55 bp 5' of the expected capped nucleotide and extending 171 bp 3' of the expected poly(A)-addition nucleotide. Clone Ptr Ch32-8.1 does not contain the complete $\Lambda\gamma$ gene, and for this reason the 3' sequence extending from the *EcoRI* site in exon 3 to 171 bp 3' of the poly(A)-addition nucleotide was obtained from lambda clone Ptr Ch30-12.6, which contains a complete $\Lambda\gamma$ gene (see fig. 1). That these two $\Lambda\gamma$ gene-containing clones represent different isolates of the same $\Lambda\gamma$ gene was established by sequencing intron 2 from both (the region expected to vary the most). These intron 2 sequences were found to be identical—even across the hot spot.

The nucleotide sequences for the chimpanzee $G\gamma$ and $\Lambda\gamma$ genes are aligned in figure 2, along with sequences of the corresponding human (Slightom et al. 1980) and gorilla (Scott et al. 1984) genes. The fetal globin genes from these hominines share a high degree of identity, not only in coding regions (exons) but also in all noncoding regions. The sequences shown in figure 2, excluding the hot spot sequence (positions 1,126-1,187), differ as follows: Hsa $G\gamma$ vs. Ggo $G\gamma$, 33 = 1.9%; Hsa $G\gamma$ vs. Ptr $G\gamma$, 39 = 2.25%; Ggo $G\gamma$ vs. Ptr $G\gamma$, 44 = 2.6%; Hsa $A\Lambda\gamma$ vs. Ggo $\Lambda\gamma$, 40 = 2.4%; Hsa $A\Lambda\gamma$ vs. Ptr $\Lambda\gamma$, 32 = 1.9%; Hsa $B\Lambda\gamma$ vs. Ggo $\Lambda\gamma$, 36 = 2.2%; Hsa $B\Lambda\gamma$ vs. Ptr $\Lambda\gamma$, 25 = 1.4%;

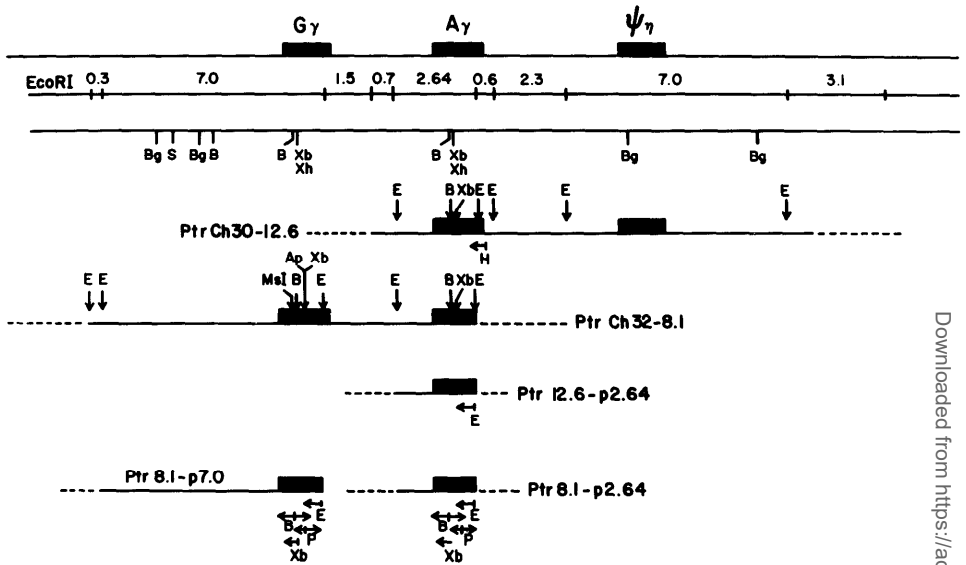


FIG. 1.—Restriction-enzyme site mapping and strategy used to sequence the chimpanzee fetal globin genes. The top line shows the organization of the human $G\gamma$ -, $A\gamma$ -, and $\psi\gamma$ -globin genes; raised bars denote the location of gene regions. The locations of various restriction-enzyme sites found in this region of the β -globin gene cluster are shown on two lines directly below (Slightom et al. 1980; Grosveld et al. 1981). Insert regions from chimpanzee recombinant lambda clones $Ptr\ Ch30-12.6$ and $Ptr\ Ch32-8.1$ have been mapped and are shown below corresponding regions of the human globin gene cluster. Most restriction-enzyme sites are shared between the human and chimpanzee β -globin gene regions. Noteworthy differences are the absence of an $EcoRI$ site in the γ intergenic region, absence of $XhoI$ and $AvaI$ sites in intron 2 of both $G\gamma$ and $A\gamma$ genes, and the presence of $MspI$ and $Apal$ sites in the chimpanzee $G\gamma$ gene. The DNA regions subcloned into pBR322 are shown below the lambda maps. Plasmid clones are $G\gamma$, $Ptr\ 8.1-p7.0$, and $A\gamma$, $Ptr\ 8.1-p2.64$ and $Ptr\ 12.6-p2.64$. Plasmid clones were used to facilitate DNA sequencing. Horizontal arrows shown below lambda and plasmid clones denote the strategy used to sequence both chimpanzee genes. Sequences were determined from both $A\gamma$ clone intron 2 regions, which were found to be identical, indicating that both lambda clones are from the same chimpanzee chromosome. Restriction enzymes are: Ap = *Apal*; B = *BamHI*; Bg = *BglII*; E = *EcoRI*; H = *HindIII*; Msl = *MspI*; P = *PvuII*; S = *SmaI*; Xb = *XbaI*; and Xh = *XhoI*.

and Ggo $A\gamma$ vs. $Ptr\ A\gamma$, 28 = 1.7%. In counting differences we scored each nucleotide substitution as well as each insertion or deletion, regardless of length, as one difference. The stretch of simple-sequence DNA $(TG)_n$ is located in the expected region of each chimpanzee γ gene and has a value of $n = 22$ for the $G\gamma$ and 14 for the $A\gamma$ gene (fig. 2). Because of gene conversions (see below), the percentage difference values for $G\gamma$ -to- $G\gamma$ and $A\gamma$ -to- $A\gamma$ comparisons among these species are larger than would be expected for comparisons of strictly orthologous genes over their full alignment; i.e. some of the differences among the 1,843 aligned positions are due to paralogous exchanges between $G\gamma$ and $A\gamma$ sequences that predate the separations of *Pan*, *Homo*, and *Gorilla*.

The 5' noncoding regions share identical promoter elements (AATAAA) located 31 bp 5' of the expected capped nucleotide, and the 3' noncoding regions share the same terminator codons and poly(A)-addition signal (AATAAA). Intron 1 sequences for all these genes are 122 bp in length, while intron 2 sequences vary in length, the chimpanzee's $G\gamma$ having 896 bp and its $A\gamma$ having 887 bp in intron 2.

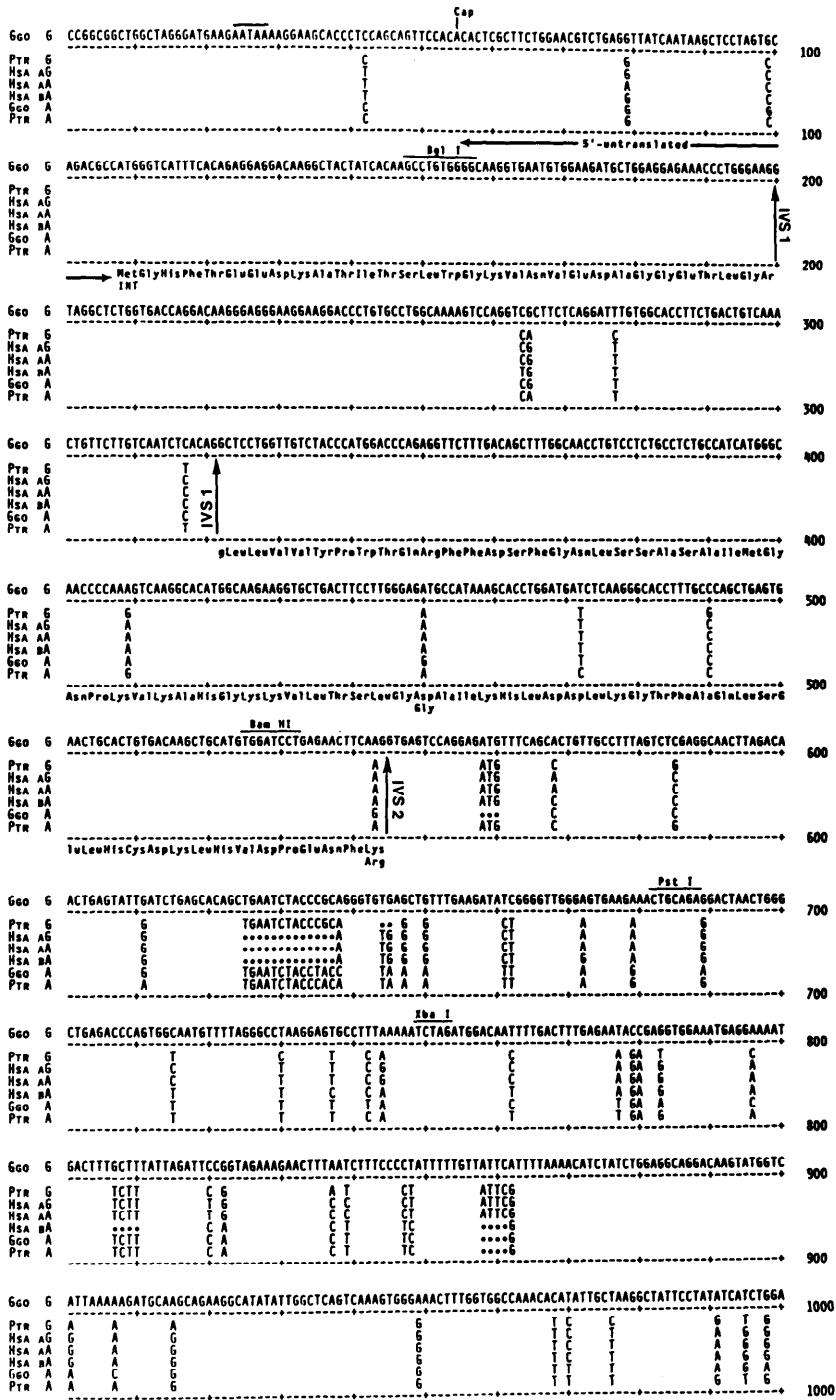


FIG. 2.—Nucleotide sequence of chimpanzee $G\gamma$ and $A\gamma$ fetal globin genes aligned with fetal-globin-gene sequences from man and the gorilla. Human $G\gamma$ - and $A\gamma$ -globin gene nucleotide sequences are from Slightom et al. (1980) and Shen et al. (1981) and are denoted as Hsa aG and Hsa aA ($G\gamma$ and $A\gamma$ genes, respectively) from human chromosome A and Hsa bA ($A\gamma$ gene) from chromosome B. Gorilla $G\gamma$ - and $A\gamma$ -globin gene nucleotide sequences are from Scott et al. (1984). The numbering system was set by the overall alignment of these gene sequences, and asterisks indicate gaps used to increase sequence identities among

Downloaded from https://academic.oup.com/mbe/article/2/5/370/974044 by guest on 21 August 2022

Comparison of the coding sequences from both chimpanzee $^G\gamma$ and $^A\gamma$ genes with those from the corresponding human and gorilla genes shows very few substitutions, either silent or replacement, among the 876 coding base positions. Substitutions that resulted in replacement of two amino acids in the gorilla $^A\gamma$ gene have been discussed by Scott et al. (1984). No substitutions causing amino acid replacements were found between the human and chimpanzee coding regions, an expected result because the amino acid sequences of chimpanzee fetal globin chains showed no difference from the human fetal globin chains (DeJong 1971). However, we find five silent-site differences between the human and chimpanzee coding regions, three in $^G\gamma$ at codons 59, 86, and 118 and two in $^A\gamma$ at codons 59 and 80 (fig. 2).

Rooted Parsimony Trees of Different Sequence Regions

Inasmuch as the rabbit γ globin gene shows a substantial number of base matches with the hominine γ genes (as illustrated by the matches with gorilla $^A\gamma$ in fig. 3), it was possible to construct rooted parsimony trees depicting the evolutionary history of different sequence regions where gene conversions occurred (figs. 4 and 5). These trees, in particular those for the exons or coding regions (fig. 5), indicate that the five silent-site differences result from four silent base substitutions in the descent of the chimpanzee genes and from a substitution in codon 59 of a protochimpanzee $^A\gamma$ gene spreading by gene conversion to the chimpanzee $^G\gamma$ gene (figs. 4 and 5).

When $^A\gamma$ and $^G\gamma$ sequences of man, the chimpanzee, and the gorilla are broken down into regions in which convergence is minimized (as in fig. 4), 15.3% fewer mutational events are required than if the sequences were treated in their entirety (i.e., a common history found for all regions). Even with these alternative arrangements for species relationships that yielded less parsimonious trees, the positioning of converted regions did not change, even though the direction of conversions did. As indicated in the Material and Methods section, regional boundaries that minimized the incidence of convergent mutations within the circumscribed regions could be placed at gaps and areas of high mismatch because such nucleotide divergences tend to terminate conversions.

Gene Conversions between Nonallelic Fetal Globin Genes

Comparisons among the $^G\gamma$ and $^A\gamma$ fetal globin genes from man and the gorilla have provided evidence for the occurrence of conversion event C1 between the two nonallelic γ loci before divergence of *Homo* and *Gorilla* but long after the original duplication in the early catarrhines (Scott et al. 1984). Two derived amino acid replacements yielding His in codon 77 and Thr in codon 135 are shared by both $^G\gamma$ and $^A\gamma$ chains of man, the chimpanzee, and the gorilla but are absent from the old-world monkey and rabbit γ chains. That is, the two tandem loci in all Hominae are cladistically closer to one another than to any nonhominoid γ gene. Given the silent-substitution rate typical for coding sequences of mammalian β -type globin genes (Efstratidis et al. 1980), we would expect at least 28 silent mutational events between any 5' γ and 3' γ gene pair if no gene conversion had occurred between them since the time of the original duplication (35 Myr ago for basal catarrhines). However, since the three hominine $^G\gamma$ and four hominine $^A\gamma$ genes last shared a common ancestral sequence, only four silent substitutions have occurred (two within $^G\gamma$, one in $^A\gamma$, and one common to $^G\gamma$ and $^A\gamma$ genes of chimpanzee; see figs. 2 and 5). Further evidence for C1 comes from comparison of the most rapidly evolving γ -gene region among

```

Rab Y GGGCCAGATGTCACGACGGAGAAGAATAAAAAGCAGCGCTTAGAGCAGTTCACATACTTCTCTGAGACATCTGAGACATACAGCAAGCTCACGGC 100
Gor Y CCGG CGGCTGGCTAGGGATGAAGAATAAAAAGGAAGCACCCTCCACAGCTTCCACACACTCGCTCTGGAAAGCTGAGGTTATCAATAAGCTC
*
Rab Y                               INTExon 1-----
Rab Y CTCTAGCACAGACATGATGGTCATTTCACCGGCTGAGGAAAGGCTGCTATCACAGACACATGGAAGCTGGTGGAGCTAGAAGATGCCGGAGCAGGAAGC 200
Gor Y CTAGTGCAGACGCCATGGTCATTTCACAGAGGAGGACAAGGCTACTATCACAAAGCTGTGGGCAAGGTGAATGTGAAGATGCTGGAGGAGAAAC
----->
Rab Y CCTCGAAAGGTAGGCCCTGGGCTCCAGGACAAGGACAGAGGAAAGTAAAGTGGAGCTGCAGGACTCCAGGCCCTTCTCAGGACTGT GAAGCTCTC 300
Gor Y CCTCGAAAGGTAGGCCCTGGT GACCAAGACAAGGAGGGAAGGAAGGACCTGTGCCTGCCAAAGTCCAGGTGCTTCTCAGGATTTGTGGACACT TC
*
Rab Y                               Exon 2-----
Rab Y TGATACTCCACCACCTAATGTTC TGTC TCT ATAGGCTCTGGTCTGTACTCTTGAGCCACAGAGATCTTCGATAGCTTTGGAACACTGCTCTCTT 400
Gor Y TG ACTGTCAA  ACTGTCTTGTCAACTCTCACAGGCTCCTGGTGTCTACCCATGGACCAGAGGTTCTTGACAGCTTTGGCAACTGCTCCTCG
-----
Rab Y CCTCTGCCATCATGGGAAACCCCAAAGTGAAGGCCATGCCAAGAAGGTGTTGACTGCCTTTGGAGATGCTGAAGAATGTGGATGATCTCAAACAC 500
Gor Y CCTCTGCCATCATGGGCAACCCCAAAGTCAAGGCACATGCCAAGAAGGTGCTGACTTCTTGGGAGTGCCATAAGACACCTGGATGATCTCAAGGGCAC
-----
Rab Y ATTGCCACCCTGAGTGARCTGCACTGTGACAGGCTGCATGGATCCCGAGAAGTCTTCAAAGTAAAGTTCAAGACAAGCTCAAGGCCCTTGCTGATTTAGTC 600
Gor Y CTTTGCCACAGCTGAGTGAAGTGCATGTGACAGCTGCATGTGGATCCTGAGAAGCTTCAAGGTTGAGTCCAGGAG TTTCAGCCCTTGGCTTTAGTC
-----
Rab Y ATGGGTGAAGTGAAGC ATTAAGTATGGCTTCTGACGACATG ***** GAGATATT TGGTGGAGATGAA 700
Gor Y TCAGGCAACTTA GACAACAGTAGATTG ATCG AGCACAGCTGAATCTACTACCGGGTGAAACTATTGAAGATATTGGGTTGGAGTGAAG
-----
Rab Y AAAGCTCTAGAGG AACTGAACTGAGACTAAGTGGTAATGCATTAGGCTTAAACAAT ACTCAGGAGTGAGAGGATTCGCTATACTCAAGTB 800
Gor Y GAA CTCGAGAAGACTAAGTGGCTGAGACCAGTGGTAAATGTTTAGGG CTAAGGAGTGCTTTAAAAATCTAGTGA CAACT TTG
-----
Rab Y TATTTTAGAAGAG GCTTCTGTAGAGATGAGGAACAGACTTTTGTTAAATGACTTTTGAAGAAATAACTCTTCATTTCTCCCAACTCAGCTTTCCT 900
Gor Y ACTTTGAGATAGAG AAGTGGAAATGAGGCAATGACTTTTCTTATTAGATCCAGTGAAGAAGACT TCACTCTTCC CACTTTTGT
-----
Rab Y TAAAAAG TTTAAAC CTA ATGAAGCA GAAAAGG A GT GACA ATG GG AAAG ATAAATGAGGGCAGCCATAA GGCAGAA 1000
Gor Y T GTTTAAACACTATCTGAGGACGAGCAAGTAGTGGCTAATTAAACAGATGAGGCAAGGCATATATTGCTCAGTCA AAGTGGGA A
-----
Rab Y CTTACGTGATTTGTGTAGAGTGCCATGCATATTCCTTAATGGCTA ATACTGTTTAAAATGCTG CAATGC CCATTTCAACCTAAATTTATACACT 1100
Gor Y CTTTGGTGGCCAAACATATATTGCTAAGGCTATTCTTATATCAGCTAGACAC ATATAAAATGCTGCCAATGCTCAACTACAACTTATATCTTAT
-----
Rab Y TCCAGAT TGACCATTATG TGTCTCTGGAGA GAA YATTGGAACATTTGGTCCGAAGT GGTTTTGAAG AGAT 1200
Gor Y TCCAGATGGGGCA AAGCATGTCCAGGGGTGAGGAACAATTGAAACATTTGGGCTGGATAGATTTTGAAGTCAAGCTGTTGTGTTGTGTGTGTT
-----
Rab Y ACCGTGAGGTTTACCTTTCC CCTCAGCCACAAA ATGGGGAGTT GTGTTAG ATGATGACAGAAATATTTAAATATGACCAAT 1300
Gor Y GTGTGTGTGCAAGCT GTGTTCTTTTAACTCTTCAOCTCAACAATACAGGGTTCATGTTGGCAAGAAATAGCAAGATTTAAATATGCCAGT
-----
Rab Y GATTAGT TGGACAGGGGACCAAACTACTACATGTAAACAAGCTGGAGGGGGGCAAGA TCAGCCTTGGAGTAAATTACAAAGGCTTGGTTTGGAG 1400
Gor Y GACTAGTGTGCA AGAAGAAC AACTACTTGCAATTTAA TGGGA AGGCAAAATCTCAGGCTTAGAGGGAAGTAAACATAGGCTGATCTGGG
-----
Rab Y GGGTA TTCAGTCCAGATAGCGGGGAACTAGCCT GAACTCTCTAATTAACATATGTTTGTGTTGTTCTCTCTTTTCACTTCACAGCTCTTAGGAA 1500
Gor Y YGGAAGCTGGGTGATGATATCTGGAGGCCAGGCTGAGACTCTCAAGCTCACTATGGTTCATCTTTATGTTCTCTTTTCACTCAACAGCTCTGGAA
-----
Rab Y ATGTGCTAGTAATTTGTCTGGCAGATTTTGGCAAGGAACTCACTCCCAAGTCCAGAGTCCCTGCCAGAAATTTGGTGGCTGGAGTGGCCACTGCCCT 1600
Gor Y ATGTGCTGGTACCCTTTGGCAATCAATTCGGCAAGAAATCAACCCCTGAGGTGCAGGCTCTCGGCAAGAAAGTGGTGAAGTGCCAGTGGCCACTGCCCT
-----
Rab Y GGGCCACAGTATCACTGAACCCCT CACTGTAGGA CAGAGC TTCTAGGA AGAAGCTTATCCCTCAAATAAATGAAATAATAAACTA C 1700
Gor Y GTCCCTCAGATACCCTGAGCTCTTGCC ATGATTACAGAGCTTCAAGGATAG GCTTATTCTGCAAGCAATA CAATAATAAATCTTTC
-----
Rab Y TCTAAGAATAATTTGTGATGGTATTGAG TTTATTTCTTGACTTTTTAAATATAGGCTCAAGGAAATTTATTTTTTATGTTGGAAGGA 1800
Gor Y TGTGGAGAGATCACACATGATTTCTCAGCTCTTTTTTACATCT TTTAAATACATGAGCCACAAGGGTITA AATTG AGGGA
-----
Rab Y TGTGTGTGTGTTGTTGTG TGCCTG TGTACTATG GTGTATC TCTCATTCTCTTTGGCTGAGCGGTGATTTCTGATGAGAAAA 1900
Gor Y AGTGTATGATGATTTCTGCACTGCTTGTG GTTGTGGGTGTCATGCTGACTTATTTTTATATGAGATGCAATT TGTAGCAAAA

```

Fig. 3.—DNA sequence comparison of rabbit γ and gorilla γ -hemoglobin genes. The rabbit sequence is from Hardison (1981), and the gorilla sequence from Scott et al. (1984). Blank spaces within these two sequences are gaps used to maximize identities. All gaps fall in the noncoding regions and were found by the algorithm of Smith and Waterman (1981), as modified by B.F.K. Vertical bars between the rabbit and gorilla sequences designate base matches. Asterisks above the rabbit sequence designate positions that vary among the hominine genes, and the symbol (V) indicates a gap shared by rabbit γ and gorilla γ but not by other γ gene sequences.

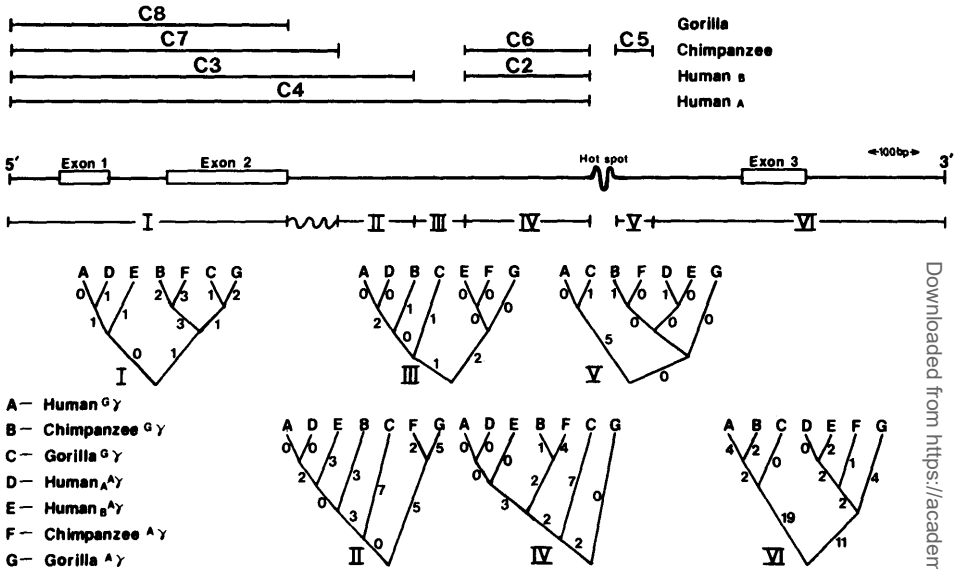
Regions of high intraspecific similarity between $G\gamma$ and $A\gamma$ 

FIG. 4.—Division of the 1,843 aligned sequence positions of hominine $G\gamma$ and $A\gamma$ globin genes into six regions, each with a different evolutionary history as depicted by the maximum parsimony method. The alignment regions and the parsimony trees describing the different evolutionary histories are diagrammed in mid and lower latitudinal portions of the figure, respectively. These alignment regions, in turn, were determined by those regions (upper latitudinal portion of figure) of high intraspecific sequence similarity between $G\gamma$ and $A\gamma$ genes that appear to have resulted from gene conversions. The separate gene conversion events (C2–C8) are assigned for the region of the particular hominine chromosome involved (see text).

hominine species, position 636–861 (in the 5' third of intron 2). We find (table 1) that in this region the paralogous divergence between $G\gamma$ and $A\gamma$ genes is only 1.5 times larger than the average orthologous divergence among human, chimpanzee, and gorilla $G\gamma$ genes and between chimpanzee and gorilla $A\gamma$ genes. The human $A\gamma$ genes were not used for this comparison because this region has been converted either totally, as in $A^A\gamma$, or partially, as in $B^A\gamma$, by human $G\gamma$ sequences. In contrast, the sequence region 3' of exon 3 (positions 1,582–1,843) shows a smaller degree of orthologous divergence but a much larger degree of paralogous divergence, ~14%–15% (table 1). This downstream region of the duplicated loci may have escaped all gene conversion events (Shen et al. 1981; Scott et al. 1984).

The rapidly evolving intron 2 sequences, upstream of the hot spot, not only provide evidence for C1 but also for conversion events C2, C3, and C4 (Scott et al. 1984), which occurred in protohuman γ genes after *Homo*, *Pan*, and *Gorilla* last shared a common ancestor. The genealogical relationships depicted by the parsimonious tree constructed for the hominine γ genes over alignment positions 636–811 (region II in fig. 4) show that the human $A^A\gamma$ and $B^A\gamma$ sequences each have been converted into $G\gamma$ sequences. The conversion of $A^A\gamma$ occurred sometime after that of $B^A\gamma$. These deductions follow because the $A^G\gamma$, $A^A\gamma$, and $B^A\gamma$ human branch is cladistically closest to chimpanzee $G\gamma$ and because human $A^G\gamma$ and $A^A\gamma$ share two derived substitutions not shared by human $B^A\gamma$ and chimpanzee $G\gamma$. Indeed, from alignment position 1 to the 5' beginning of the hot spot, i.e., position 1,128, human $A^A\gamma$ always

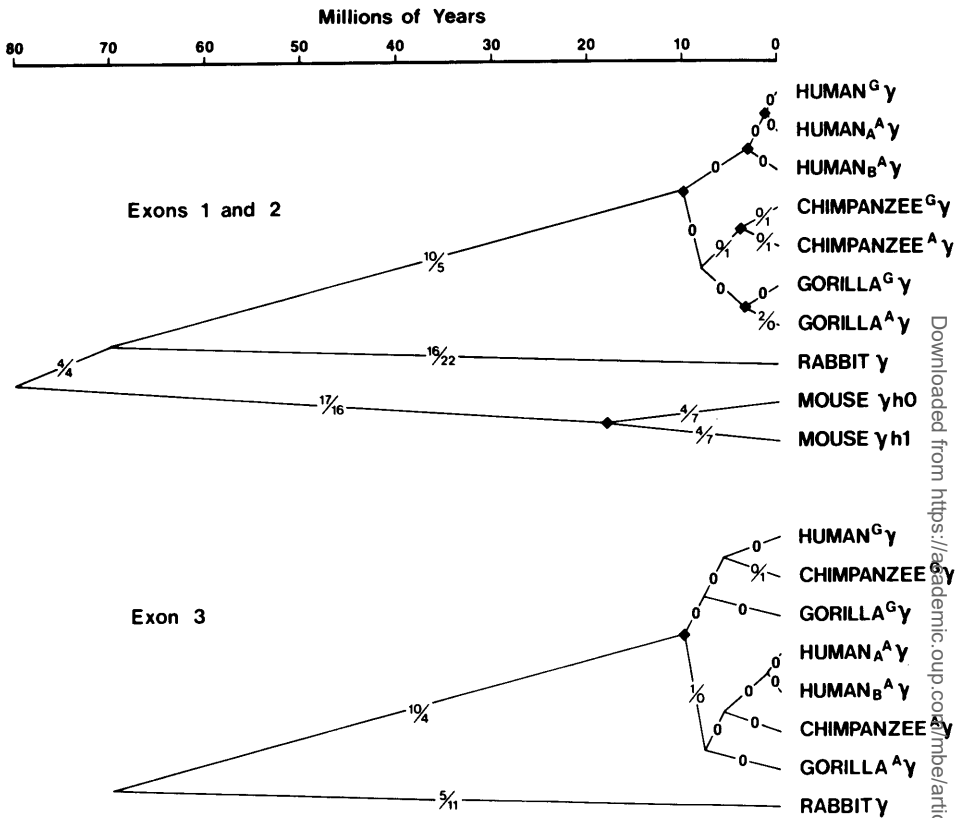


FIG. 5.—The γ -hemoglobin branches of the phylogenetic tree constructed by the maximum parsimony method for coding sequences representing 64 globin genes and pseudogenes. As depicted the sequences spanning exons 1 + 2 have a different history from that of the sequences spanning exon 3. Link lengths, calculated by the method described in Czelusniak et al. (1982), are shown as fractions, with the numerator as the number of amino acid-changing (nonsynonymous) nucleotide substitutions and the denominator as the number of silent (synonymous) substitutions. Nodal points representing paralogous separations are designated by black diamonds. See text for the rationale used to place the nodal points on the ordinal time scale of millions of years.

groups first with human $A^G\gamma$ (see trees *I-IV* in fig. 4), while in three of these regions (*II*, *III*, and *IV*) the chimpanzee and gorilla $G^G\gamma$ sequences are both orthologically related to human $G^G\gamma$. This grouping supports our earlier conclusion (Slightom et al. 1980; Scott et al. 1984) that in the conversion event labeled C4, the $A^G\gamma$ sequence in the human $A^G\gamma$ locus was replaced by the $G^G\gamma$ sequence.

The conversion event or events affecting the human $B^A\gamma$ locus, labeled C2 and C3 (Scott et al. 1984), occurred in two separate regions of the alignment. C2 spans position 862–1,128, where the $B^A\gamma$ locus (like $A^G\gamma$ in the case of C4) has a $G^G\gamma$ sequence (*IV* in fig. 4), and C3 spans positions 1–811 (*I* and *II* in fig. 4), where, again, the $B^A\gamma$ locus contains a $G^G\gamma$ sequence. C2 and C3 could be the result of a single event rather than the two conversions that we have suggested. A single conversion of the proto- $B^A\gamma$ locus in which strands in the heteroduplex looped out, skipping positions 812–861, can be hypothesized to have occurred. Such a conversion mechanism has been referred to as a bubble or patchy conversion (Kourilsky 1983; Stoeckert et al. 1984).

Table 1
Representative Orthologous and Paralogous Pairwise Comparisons among Hominine $G\gamma$ - and $A\gamma$ -Hemoglobin Genes in Two Sequence Regions

Region and Type of Comparison	Nucleotide Differences \pm Gaps*/ Shared Positions (%)
Intron 2:†	
<i>Orthologous</i> (636–861):	
Chimpanzee $G\gamma$ vs. human $G\gamma$	8/222 (3.6)
Chimpanzee $G\gamma$ vs. gorilla $G\gamma$	13/224 (5.8)
Human $G\gamma$ vs. gorilla $G\gamma$	12/224 (5.5)
Chimpanzee $A\gamma$ vs. gorilla $A\gamma$	12/220 (5.5)
<i>Paralogous</i> (636–861):	
Chimpanzee $G\gamma$ vs. chimpanzee $A\gamma$	16/220 (7.3)
Chimpanzee $G\gamma$ vs. gorilla $A\gamma$	17/220 (7.7)
Gorilla $G\gamma$ vs. gorilla $A\gamma$	20/220 (9.0)
Human $G\gamma$ vs. chimpanzee $A\gamma$	16/220 (7.3)
Human $G\gamma$ vs. gorilla $A\gamma$	20/220 (9.0)
3' of exon 3:	
<i>Orthologous</i> (1,582–1,843):	
Chimpanzee $G\gamma$ vs. human $G\gamma$	5/256 (2.0)
Chimpanzee $G\gamma$ vs. gorilla $G\gamma$	3/256 (1.2)
Human $G\gamma$ vs. gorilla $G\gamma$	6/256 (2.3)
Chimpanzee $A\gamma$ vs. human $A\gamma$	1/255 (0.4)
Chimpanzee $A\gamma$ vs. gorilla $A\gamma$	3/252 (1.2)
Human $A\gamma$ vs. gorilla $A\gamma$	4/252 (1.6)
<i>Paralogous</i> (1,582–1,843):	
Chimpanzee $G\gamma$ vs. chimpanzee $A\gamma$	36/253 (14.2)
Chimpanzee $G\gamma$ vs. human $A\gamma$	35/253 (13.8)
Chimpanzee $G\gamma$ vs. gorilla $A\gamma$	38/250 (15.2)
Human $G\gamma$ vs. human $A\gamma$	34/255 (13.3)
Human $G\gamma$ vs. chimpanzee $A\gamma$	35/255 (13.7)
Human $G\gamma$ vs. gorilla $A\gamma$	38/252 (15.1)
Gorilla $G\gamma$ vs. gorilla $A\gamma$	35/250 (14.0)
Gorilla $G\gamma$ vs. chimpanzee $A\gamma$	33/253 (13.0)
Gorilla $G\gamma$ vs. human $A\gamma$	32/253 (12.6)

* Each gap counts as one difference regardless of its length.

† Rapidly evolving.

However, to account for the fact that over positions 862–1,128 human $B\gamma$ is identical to both human $A\gamma$ and human $A\gamma$ and also shares three substitutions not found in chimpanzee $G\gamma$ (see *IV* in fig. 4), we must assume that this region became very conservative in both $G\gamma$ and $A\gamma$ loci after the bubble-conversion event. Alternatively, we can suggest that conversion C2 be retained but have its sequence derived either from an allelic $A\gamma$ gene that has already been converted in this region by the $G\gamma$ sequence or from a nonallelic conversion involving a conserved donor sequence from a $G\gamma$ locus.

In addition to providing additional support for gene conversion events in the evolution of the human γ loci, the parsimony trees in figure 4 indicate that separate gene conversions have occurred in the descent of chimpanzee and gorilla γ loci. The tree for region *IV* (spanning alignment positions 862–1,128) depicts not only the human C2 conversion but also an independent parallel conversion (labeled conversion C6 in fig. 4) between chimpanzee $G\gamma$ and $A\gamma$ loci in which an $A\gamma$ sequence was replaced

by a $G\gamma$ sequence. Note that in this tree chimpanzee $G\gamma$ and $A\gamma$ are cladistically closer to one another than to any human or gorilla genes and also that the enlarged branch of three human and two chimpanzee genes is cladistically closer to gorilla $G\gamma$ than to gorilla $A\gamma$.

Immediately downstream from the hot spot (starting at position 1,188 at the 3' end of the hot spot and extending downstream through position 1,254), the chimpanzee $G\gamma$ gene is almost identical to $A\gamma$ genes of the gorilla, man, and the chimpanzee, whereas human and gorilla $G\gamma$ genes share five derived substitutions not found in the $A\gamma$ genes (figs. 2 and 4). We conclude that in the chimpanzee $G\gamma$ locus, the $G\gamma$ sequence has been replaced by the $A\gamma$ sequence (labeled conversion C5 in fig. 4). Because the $A\gamma$ hominine sequences in this region prove to be very conservative (e.g., at positions where they vary from human and gorilla $G\gamma$ sequences, they share the same nucleotides with rabbit γ), we cannot conclude whether the postulated conversion was recent or not. It could have occurred almost as anciently as the ancestral divergence of *Pan* from *Homo*.

Aside from these two conversions extending from the hot spot (one upstream and the other downstream), another conversion (labeled conversion C7 in fig. 4) between $G\gamma$ and $A\gamma$ chimpanzee loci is evident in a distant upstream region spanning positions 1–635, which encompasses exon 1, intron 1, exon 2, and the first 90 bases of intron 2. The parsimony tree constructed for this region (*I* in fig. 4) has chimpanzee $G\gamma$ and $A\gamma$ genes cladistically closer to one another than to any human or gorilla genes. Similarly, over much of this region (positions 1–557) a conversion may have occurred between the two nonallelic gorilla genes (labeled conversion C8 in fig. 4). Note that for region *I* the gorilla $G\gamma$ and $A\gamma$ genes are depicted as being cladistically closer to one another than to any human or chimpanzee genes (see *I* in fig. 4). On comparing the branching pattern of region *I* tree to trees for other regions, we can deduce that region *I* for the two chimpanzee and two gorilla genes consists of $A\gamma$ sequences, while region *I* for the three human genes consists of $G\gamma$ sequences. In region *I* the chimpanzee branch ($G\gamma$ plus $A\gamma$) and gorilla branch (again, $G\gamma$ plus $A\gamma$) group together first before joining the human branch. Yet, in comparisons of orthologous sequences, *Pan* and *Homo* group together first before being joined by *Gorilla* (see *II*, *IV*, and *VI* in fig. 4). Moreover, because in region *II* immediately downstream from region *I* our evidence indicated that $G\gamma$ sequences replaced $A\gamma$ sequences in human *A* and *B $A\gamma$* loci, it is likely (and more parsimonious) that this is also the case in region *I*. Thus in region *I*, all three human genes (having $G\gamma$ sequences in region *I*) are paralogously related to all chimpanzee and gorilla genes (having $A\gamma$ sequences in region *I*).

Although we may treat the three discontinuous regions where gene conversions are evident between chimpanzee $G\gamma$ and $A\gamma$ loci as being the result of three separate events, we might also account for these conversion results by a single event, using the random patchy conversion model (Stoekert et al. 1984). Such a single event would have the $G\gamma$ locus of an ancestral chimpanzee converted to the $A\gamma$ sequence over positions 1–635 and 1,188–1,254 while at the same time the $A\gamma$ locus accepted $G\gamma$ sequence over positions 862–1,128. The remaining regions (636–861 and 1,255–1,848) of this $A\gamma$ locus would remain unconverted. Michelson and Orkin (1983) have found a patchy pattern of conversion regions between two nonallelic human α -hemoglobin genes, suggesting that each heteroduplex strand served as a template for mismatch repair that resulted in a mosaic arrangement of parental sequences over the conversion regions.

Role of Hot Spot Sequence in Fetal-Globin-Gene Conversions

The large intron of all hominine γ -globin genes sequenced thus far (those of man, gorillas, and chimpanzees) contain a simple-sequence DNA, consisting of polypurine-pyrimidine bases (see alignment positions 1,127–1,187 in fig. 2). It has been argued that because polypurine-pyrimidine DNA stretches tend to form an alternative structure of DNA, known as Z-DNA (Wells et al. 1982), they might be more accessible to nuclease activity (Shen et al. 1981). Kilpatrick et al. (1984) showed that all three hot spot sequences contained in the human genes shown in figure 2 can form Z-DNA structures and that the boundary regions between B- and Z-DNA conformations are sensitive to S1 nuclease digestions. The Z-DNA conformation not only provides access for nucleases, but, as Kilpatrick et al. (1984) suggest, the nonhelical regions at the junction between B- and Z-DNA may serve as entry points for strand invasions. Thus it appears that the Z-DNA conformation may indeed increase strand cleavages and transfers, which are the first two steps necessary to initiate either gene conversion or crossover (Radding 1982). The fact that four of the seven regions of high intraspecific similarity between $G\gamma$ and $A\gamma$ loci (shown in fig. 4) are bordered by the hot spot supports this hypothesis. However, the actual region of a heteroduplex strand used as a template for mismatch repair need not necessarily be close to the site of strand cleavage and transfer. For example, the postulated gorilla conversion (see region I in fig. 4) encompassed exons 1 and 2 and intron 1 but terminated in the far 5' side of intron 2, long before reaching the hot spot. We cannot rule out the possibility that strand cleavage and transfer leading to a heteroduplex was initiated by the hot spot; however, the actual repair template employed was limited to the highly conserved 5' region of these γ -globin loci.

In fact, the overall evidence on gene conversions, including that from studies on fungi (Nicolas and Rossignol 1983) as well as from studies on mammalian globin genes (Schon et al. 1982; Michelson and Orkin 1983; Goodman et al. 1984; Hardies et al. 1984), indicates that conserved DNA sequences such as coding sequences engage in gene conversion and crossing-over more often than do variable sequences such as introns. Conversions of δ by β sequences involved exons 1 and 2, intron 1, and the conserved region 5' of exon 1 up to and including the CCAAT box but not intron 2 and the 3' noncoding flanking region (Goodman et al. 1984; Hardies et al. 1984). The crossover boundary for the hybrid 5'- $\psi\eta\delta$ -3' gene of the lemur is at the 3' end of exon 2. Similarly, mouse $\beta h0$ - and $\beta h1$ -globin genes are 5'- $\gamma\epsilon$ -3' hybrids with the boundary between γ and ϵ occurring at the 5' start of intron 2. The hot spot sequence in intron 2 of the hominine γ genes, therefore, is probably not necessary for gene conversion, but it may provide a stimulus for gene conversion.

Branching Pattern within Homininae

We have already indicated that, of gene regions where chimpanzee $G\gamma$ sequences appear to be orthologously related to human and gorilla $G\gamma$ sequences, the chimpanzee and human sequences grouped together first before the gorilla sequence joined them. Note again in fig. 4 the parsimonious trees constructed for regions II, IV, and VI. In the most extensive stretch of sequence (positions 1,254–1,843) where the chimpanzee $A\gamma$ sequence is orthologously related to human and gorilla $A\gamma$ sequences, the chimpanzee and human $A\gamma$ sequences are closest cladistically. These findings agree with the parsimony evidence from α -hemoglobin amino acid sequences (Goodman et al. 1983)

and $\psi\eta$ -globin nucleotide sequences (Goodman et al. 1984) as well as being suggested by cellular DNA-hybridization studies (Hoyer et al. 1972; Sibley and Ahlquist 1984). However, it is only slightly less parsimonious to first group *Pan* and *Gorilla* in these several sets of sequence data. With regard to the present data on γ -hemoglobin genes, grouping *Homo* and *Pan* first requires four fewer substitutions than grouping *Pan* and *Gorilla* first and nine fewer substitutions than if *Homo* and *Gorilla* are joined first. These scores result from the finding that there are 10 alignment positions (661, 742, 777, 970, 1,018, 1,067, 1,602, and 1,816 for $G\gamma$ sequences and 1,291 and 1,292 for $A\gamma$ sequences; see fig. 2) where the lowest substitution score is obtained by grouping *Homo* and *Pan* first. This compares to six alignment positions (837, 915, 995, 1,640, and 1,699 for $G\gamma$ sequences and 1,715 for $A\gamma$ sequences) where *Pan* must be grouped first with *Gorilla* for the lowest substitution score and only one position (715 for $G\gamma$ sequences) where the lowest score is obtained by first grouping *Homo* and *Gorilla*.

Clearly, the present set of γ -hemoglobin sequences are far from optimal for resolving the *Homo/Pan/Gorilla* trichotomy because there are no data available for an outgroup closer to Homininae than the rabbit $\gamma:\beta 3$ -hemoglobin gene sequence. A decisive analysis may be obtained once orangutan γ -hemoglobin genes are sequenced, since *Pongo* (orangutan) is the sister taxon of Homininae (Goodman 1976; Goodman et al. 1983). The sequencing of the $\psi\eta$ -globin DNA region of the β -globin gene cluster of the orangutan should also contribute to resolving the cladistic branching pattern within Homininae, since this region has already been sequenced in *Homo*, *Pan*, and *Gorilla* (Chang and Slightom 1984).

Decelerated Rate of Coding Sequence Evolution in Hominines

Fig. 5 shows the branches of the maximum parsimony tree constructed for aligned exons of 64 sequenced globin genes. The history of exon 3 sequences is shown separately from that of exon 1 + 2 because the same gene conversions encompassing exon 2 of the hominine γ loci also encompassed exon 2 but not exon 3. Also, the mouse $\beta\text{H}0$ and $\beta\text{H}1$ loci are $\gamma\epsilon$ hybrids, with exon 1 + 2 being in the γ region and exon 3 being in the ϵ region. The numbers on the links between nodes (shown as fractions) are the number of amino acid-changing base substitutions (numerator) over the number of silent or synonymous base substitutions (denominator). Because conversions have occurred between $G\gamma$ and $A\gamma$ loci in a common ancestor of *Gorilla*, *Homo*, and *Pan*, the most ancient paralogous separation between the two nonallelic loci is taken (as in our previous study, Scott et al. [1984]) as 10 Myr ago (at the base of Homininae) rather than 35 Myr ago. Similarly, because of conversions between the two nonallelic α -hemoglobin loci of hominines, their most ancient paralogous separation (as determined from sequence data on these loci in the chimpanzee and in man) was taken as ~ 10 Myr ago (Liebhaber and Begley 1983), again at the base of Homininae. In turn, the ancestral divergence of *Pan* from *Homo* was placed in the range of 5–7 Myr ago, as suggested by molecular-clock calculations using DNA sequence data (Goodman et al. 1984; Sibley and Ahlquist 1984) and current paleontological assessments (Andrews 1982; Pilbeam 1984). The divergence of primates from Lagomorpha and other eutherian orders was placed in the range of 70–85 Myr ago, following a consensus among paleontologists and molecular evolutionists.

Using the time scale described above, our calculations show (table 2) that a marked deceleration of rates of coding-sequence evolution in globin genes occurred during descent of the hominines. We find an 18-fold decrease in the nonsynonymous (amino

Table 2
Rates of Nucleotide Substitutions in Exons of Globin Genes during Evolution from Early Eutherians to Present-Day Hominines (Chimpanzees, Humans, and Gorillas)

RATE	EVOLUTION TO THE PRESENT DAY FROM		
	Early Eutherians (α , ϵ , γ , β) (85–70 Myr)	Primates (α , γ) (70–10 Myr)	Hominines (α , γ) (10–0 Myr)
Nonsynonymous	1.8	0.9	0.1
Synonymous	6.6	1.9	0.9

NOTE.—The numbers of nonsynonymous and synonymous substitutions on the relevant links of the maximum parsimony tree for coding sequences of 64 globin genes and the time scale described in the text were employed for these calculations. The γ -hemoglobin branches of this tree are shown in fig. 5; the complete tree itself is about the same as that shown in fig. 1 in Goodman et al. (1984), except for addition of chimpanzee $\alpha\gamma$ and $\alpha\gamma$ and treatment of mouse $\beta h0$ and $\beta h1$ as $\gamma\epsilon$ hybrids. All data are expressed as multiples of 10^9 substitutions/site per year.

acid-changing) rate and a 7.5-fold decrease in the synonymous rate in the descent from early eutherians (85–70 Myr ago) to extant hominines (10–0 Myr). The nonsynonymous rate found for hominines is 0.1×10^9 substitutions/site per year, which is one-ninth the average rate (over the past 80 Myr or so) found by Li et al. (1985) for 39 mammalian genes in interordinal comparisons. It is also one-ninth the average that we find for globin genes (α , ϵ , γ , and β) during the descent of the four mammalian orders (Primates, Lagomorpha, Artiodactyla, and Rodentia) represented by sequenced hemoglobin genes. The synonymous rate found for hominines, 0.9×10^9 substitutions/site per year, is about one-fifth the rate found by Li et al. (1985) for the 39 mammalian globin genes and one-third to one-fourth the rate that we find for globin genes during the descent of the four mammalian orders.

Conclusions

In interpreting these results on the slow rate of coding-sequence evolution in hominines, a rate that is especially slow (almost 0) in the descent of the human γ - and α -globin genes, we can hypothesize, as we have in our previous paper (Goodman et al. 1984), that enhanced DNA repair mechanisms evolved in higher primates and that such repair acts, preferentially in conjunction with purifying selection, on the regions of DNA with conserved functional sequences. The gene correction mechanism that prevents emergence of new alleles could operate during genetic recombination (Ahmand et al. 1975; Whitehouse 1982). Base mispairs would be repaired as in a biased gene conversion (Dover 1982; Radding 1982), but this bias would always favor DNA resynthesis that utilized undamaged template strands from the wild-type allele rather than those from any mutant allele. The nub of our hypothesis is that gene correction (the DNA repair that prevents emergence of new alleles) and gene conversion both utilize the same molecular mechanism of strand excision and resynthesis during duplex formation and occur over conserved DNA sequence regions such as exons more frequently than they do over rapidly evolving sequences, such as are found in intron 2 of β -type globin genes. Synergy between purifying selection and DNA repair would

then account for the rate of sequence evolution being slower in exons of genes than in other DNA regions.

Acknowledgments

We are grateful to Dr. E. A. Zimmer for the gift of chimpanzee 1 DNA. We thank Timothy W. Theisen for technical assistance and Drs. Frederick R. Blattner and Oliver Smithies for the use of shared laboratory space and equipment. This study was supported by National Institutes of Health grant HD16595 (J.L.S.) and National Science Foundation grant BSR 83-07336 (M.G.). This is paper 2752 from the Laboratory of Genetics, University of Wisconsin-Madison.

LITERATURE CITED

- AHMAND, A., W. K. HOLLOMAN, and R. HOLLIDAY. 1975. Nuclease that preferentially inactivates DNA containing mismatched bases. *Nature* **258**:54-56.
- ANDREWS, P. 1982. Hominoid evolution. *Nature* **295**:185-186.
- BARRIE, P. A., A. J. JEFFREYS, and A. F. SCOTT. 1981. Evolution of the β -globin gene cluster in man and the primates. *J. Mol. Biol.* **149**:319-336.
- BENTON, W. D., and R. W. DAVIS. 1977. Screening λ gt recombinant clones by hybridization to single plaques in situ. *Science* **196**:180-182.
- BIRNBOIM, H. D., and J. DOLY. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**:1513-1523.
- BLATTNER, F. R., B. G. WILLIAMS, A. E. BLECHL, K. DENNISTON-THOMPSON, H. E. FABER, L. A. FURLONG, D. J. GRUNWOLD, D. O. KIEFER, D. D. MOORE, J. W. SCHUMNA, R. L. SHELDON, and O. SMITHIES. 1977. Charon phages: safer derivatives of bacteriophage lambda for DNA cloning. *Science* **194**:161-169.
- BREATHNACH, R. C., C. BENOIST, K. O'HARE, F. GANON, and P. CHAMBON. 1978. Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc. Natl. Acad. Sci. USA* **75**:4853-4857.
- CHANG, L.-Y. E., and J. L. SLIGHTOM. 1984. Isolation and nucleotide sequence analysis of the β -type globin pseudogene from human, gorilla and chimpanzee. *J. Mol. Biol.* **180**:767-783.
- CZELUSNIAK, J., M. GOODMAN, D. HEWETT-EMMETT, M. L. WEISS, P. J. VENTA, and R. E. TASHIAN. 1982. Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes. *Nature* **298**:297-300.
- DAGERT, M., and S. D. EHRICK. 1979. Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells. *Gene* **6**:23-28.
- DEJONG, W. W. W. 1971. Chimpanzee foetal haemoglobin: structural heterogeneity of the γ chain. *Biochem. Biophys. Acta* **251**:217-226.
- DEVEREAUX, J., P. HAEBERLI, and O. SMITHIES. 1984. A comprehensive set of sequences and analysis programs for the VAX. *Nucleic Acids Res.* **12**:387-395.
- DOVER, G. 1982. Molecular drive: a cohesive mode of species evolution. *Nature* **299**:564-572.
- EFSTRATIADIS, A., J. W. POSAKONY, T. MANIATIS, R. M. LAWN, C. O'CONNELL, R. A. SPRETZ, J. K. DERIEL, B. G. FORGET, S. M. WEISSMAN, J. L. SLIGHTOM, A. E. BLECHL, O. SMITHIES, E. F. BARALLE, C. C. SHOULDERS, and N. J. PROUDFOOT. 1980. The structure and evolution of the human β -globin gene family. *Cell* **21**:653-668.
- FRITSCH, E. F., R. M. LAWN, and T. MANIATIS. 1980. Molecular cloning and characterization of the human β -like globin gene cluster. *Cell* **19**:959-972.
- GOODMAN, M. 1976. Towards a genealogical description of the primates. Pp. 321-353 in M. GOODMAN and R. E. TASHIAN, eds. *Molecular anthropology*. Plenum, New York.
- GOODMAN, M., G. BRAUNTIZER, A. STANGL, and B. SCHRANK. 1983. Evidence on human origins from haemoglobins of African apes. *Nature* **303**:546-548.
- GOODMAN, M., B. F. KOOP, J. CZELUSNIAK, M. L. WEISS, and J. L. SLIGHTOM. 1984. The η -

- globin gene: its long evolutionary history in the β -globin gene family of mammals. *J. Mol. Biol.* **180**:803–823.
- GROSVELD, F. G., H. H. M. DAHL, E. DEBORE, and R. A. FLAVELL. 1981. Isolation of β -globin-related genes from a human cosmid library. *Gene* **13**:227–237.
- HARDIES, S. C., M. H. EDGEL, and C. A. HUTCHINSON. 1984. Evolution of the mammalian β -globin gene cluster. *J. Biol. Chem.* **259**:3748–3756.
- HARDISON, R. C. 1981. The nucleotide sequence of rabbit embryonic globin gene $\beta 3$. *J. Biol. Chem.* **256**:11780–11786.
- HARRIS, S., P. A. BARRIE, M. L. WEISS, and A. J. JEFFREYS. 1984. The primate $\psi\beta 1$ gene: an ancient β -globin pseudogene. *J. Mol. Biol.* **180**:784–802.
- HOHN, B. 1979. In vitro packaging of λ and cosmid DNA. *Methods Enzymol.* **68**:299–309.
- HOYER, B. H., M. W. VON DE VELDE, M. GOODMAN, and R. B. ROBERTS. 1972. Examination of homininae evolution by DNA sequence homology. *J. Hum. Evol.* **1**:645–649.
- HUISMAN, T. H. J., W. A. SCHROEDER, M. E. KEELING, W. GENGOZIAN, A. MILLER, A. R. BRODIE, J. R. SHELTON, and G. APELL. 1973. Search for non-allelic structural genes for γ -chains of fetal hemoglobins in some primates. *Biochem. Genet.* **10**:309–318.
- JEFFREYS, A. J. 1979. DNA sequence variants in γ , α , δ - and β -globin genes in man. *Cell* **18**:1–10.
- KAN, Y. W., and A. M. DOZY. 1978. Polymorphism of DNA sequence adjacent to minor β -globin structural gene: relationship to sickle mutation. *Proc. Natl. Acad. Sci. USA* **75**:5635–5635.
- KILPATRICK, M. W., J. KLYSIK, C. K. SINGLETON, D. A. ZARLING, T. M. JOVIN, L. H. HARRIS, B. F. ERLANGER, and R. D. WELLS. 1984. Intervening sequences in human fetal globin genes adopt left-handed Z helices. *J. Biol. Chem.* **259**:7268–7274.
- KOURILSKY, P. 1983. Genetic exchanges between partially homologous nucleotide sequences: possible implications for multigene families. *Biochimie* **65**:85–93.
- LI, W.-H., C.-C. LUO, and C.-I. WU. 1985. Evolution of DNA sequences. Accepted in R. MACINTYRE, ed. *Molecular evolutionary genetics*. Plenum, New York.
- LIEBHABER, S. A., and K. A. BEGLEY. 1983. Structural and evolutionary analysis of the two chimpanzee α -globin mRNAs. *Nucleic Acids Res.* **11**:8915–8929.
- LOENEN, W. A. M., and F. R. BLATTNER. 1983. Lambda charon vectors (Ch32, 33, 34 and 35) adopted for DNA cloning in recombination-deficient hosts. *Gene* **26**:171–179.
- MANIATIS, T., A. JEFFREY, and D. G. KLEID. 1975. Nucleotide sequence of the rightward operator of phage λ . *Proc. Natl. Acad. Sci. USA* **72**:1184–1188.
- MAXAM, A. M., and W. GILBERT. 1980. Sequencing end-labeled DNA with base specific chemical cleavages. *Methods Enzymol.* **65**:499–560.
- MICHELSON, A. M., and S. H. ORKIN. 1983. Boundaries of gene conversion within the duplicated human α -globin genes. *J. Biol. Chem.* **258**:15245–15254.
- MIESFIELD, R., M. KRYSAL, and N. ARNHEIM. 1981. A member of a new repeated sequence family which is conserved throughout eucaryotic evolution is found between the human δ and β -globin genes. *Nucleic Acids Res.* **9**:5931–5947.
- MURRAY, N. E., W. J. BRAMMER, and K. MURRAY. 1977. Lambdoid phages that simplify the recovery of in vitro recombinants. *Mol. Gen. Genet.* **150**:53–61.
- NICHOLAS, A., and S.-L. ROSSIGNOL. 1983. Gene conversion: point-mutation heterozygosities lower heteroduplex formation. *EMBO J.* **2**:2265–2270.
- PILBEAM, D. 1984. The descent of hominoids and hominids. *Sci. Am.* **250**:84–97.
- RADDING, C. 1982. Homologous pairing and strand exchange in genetic recombination. *Annu. Rev. Genet.* **16**:405–437.
- RIMM, D. L., D. HORNESS, J. KUCERA, and F. R. BLATTNER. 1980. Construction of coliphage lambda Charon vectors with *Bam*HI cloning sites. *Gene* **12**:301–309.
- SCHON, E. A., S. M. WERNKE, and J. B. LINGREL. 1982. Gene conversion of two functional goat α -globin genes preserves only minimal flanking sequences. *J. Biol. Chem.* **257**:6825–6835.

- SCHROEDER, W. A., J. R. SHELTON, J. B. SHELTON, and T. H. J. HUISMAN. 1963. The amino acid sequence of the γ -chain of human fetal hemoglobin. *Biochemistry* 2:992-1008.
- SCOTT, A. E., P. HEATH, S. TRUSKO, S. H. BOYER, W. PRASS, M. GOODMAN, J. CZELUSNIAK, L.-Y. E. CHANG, and J. L. SLIGHTOM. 1984. The sequence of the gorilla fetal globin genes: evidence for multiple gene conversions in human evolution. *Mol. Biol. Evol.* 1:371-389.
- SHEN, S., J. L. SLIGHTOM, and O. SMITHIES. 1981. A history of the human fetal globin gene duplication. *Cell* 26:191-203.
- SIBLEY, C. G., and J. E. AHLQUIST. 1984. The phylogeny of the hominoid primates, as indicated by DNA-DNA hybridization. *J. Mol. Evol.* 20:2-15.
- SLIGHTOM, J. L., A. E. BLECHL, and O. SMITHIES. 1980. Human fetal $^G\gamma$ - and $^A\gamma$ -globin genes: complete nucleotide sequence suggests that DNA can be exchanged between these duplicated genes. *Cell* 21:627-638.
- SMITH, F., and M. S. WATERMAN. 1981. Identification of common molecular subsequences. *Mol. Biol.* 147:195-197.
- STOECKERT, C. J., JR., F. S. COLLINS, and S. M. WEISSMAN. 1984. Human fetal globin DNA sequences suggest novel conversion event. *Nucleic Acids Res.* 12:4469-4479.
- WELLS, R. D., J. J. MIGLIETTA, J. KLYSKIN, J. E. LARSON, S. M. STUDIVANT, and W. ZACHARIAS. 1982. Spectroscopic studies on acetylaminofluorene-modified (dT-dG)_n-(dC-dA)_n suggest a left-handed conformation. *J. Biol. Chem.* 257:10166-10171.
- WELLER, P., A. J. JEFFREYS, V. WILSON, and A. BLANCHETOT. 1984. Organization of the human myoglobin gene. *EMBO J.* 3:439-446.
- WHITEHOUSE, H. L. K. 1982. Genetic recombination: understanding the mechanisms. Wiley, New York.
- WILSON, J. T., L. B. WILSON, J. K. DERIEL, L. VILLA-KOMAROFF, A. EFSTRATIADIS, B. G. FORGET, and S. M. WEISSMAN. 1978. Insertion of synthetic copies of human globin genes into bacterial plasmids. *Nucleic Acids Res.* 5:563-581.
- ZIMMER, E. A. 1981. Evolution of primate globin genes. Ph.D. thesis. University of California, Berkeley.

WALTER M. FITCH, reviewing editor

Received March 6, 1985; revision received May 23, 1985.