*Proc. Natl. Acad. Sci. USA*
Vol. 88, pp. 7396–7400, August 1991
Evolution

# Duplication of the γ-globin gene mediated by L1 long interspersed repetitive elements in an early ancestor of simian primates

(gene duplication/genome complexity/molecular evolution)

DAVID H. A. FITCH*†, WENDY J. BAILEY‡, DANILO A. TAGLE‡§, MORRIS GOODMAN*, LEANG SIEU¶, AND JERRY L. SLIGHTOM¶

Departments of *Anatomy and ‡Molecular Biology, Wayne State University School of Medicine, Detroit, MI 48201; and ¶Molecular Biology, Unit 7242, The Upjohn Company, Kalamazoo, MI 49007

*Communicated by James F. Crow, May 13, 1991 (received for review October 15, 1990)*

**ABSTRACT** **Regions surrounding the single γ-globin gene of galago and the duplicated $\gamma^1$- and $\gamma^2$-globin genes of gibbon, rhesus monkey, and spider monkey were sequenced and aligned with those from humans. Contrary to previous studies, spider monkey was found to have not one but two γ-globin genes, only one of which ($\gamma^2$) is functional. The reconstructed evolutionary history of the γ-globin genes and their flanking sequences traces their origin to a tandem duplication of a DNA segment ≈5.5 kilobases long that occurred before catarrhine primates (humans, apes, and Old World monkeys) diverged from platyrrhines (New World monkeys), much earlier than previously thought. This reconstructed molecular history also reveals that the duplication resulted from an unequal homologous crossover between two related L1 long interspersed repetitive elements, one upstream and one downstream of the single ancestral γ-globin gene. Perhaps facilitated by the redundancy resulting from the duplication, the γ-globin genes escaped the selective constraints of embryonically functioning genes and evolved into fetally functioning genes. This view is supported by the finding that a burst of nonsynonymous substitutions occurred in the γ-globin genes while they became restructured for fetal expression in the common ancestor of platyrrhines and catarrhines.**

Investigation of the evolutionary history of genes in the mammalian β-like globin gene family has provided insights into the processes of molecular evolution that first produced more of the same genes (duplication) and then produced, from the duplicates, diversified genes ranging from those with new functional roles to those that have become nonfunctional pseudogenes (1–7). In humans, apes, and Old World monkeys, the genes of the β-like globin gene family are clustered within a 60-kilobase (kb) genomic segment, with each gene flanked by extensive noncoding DNA, and are arranged in order of developmental expression: 5′–ε (embryonic)–$\gamma^1$ and $\gamma^2$ (fetal)–ψη (nonexpressed)–δ and β (postnatal)–3′ (2). This cluster arose through a series of gene duplications beginning 200–150 million years ago (MYA) in early mammals to produce proto-ε-globin and proto-β-globin genes (3, 6, 7). About 100–80 MYA in the stem eutheria (early placental mammals), duplications yielded δ and β from proto-β, and ε, γ, and η from proto-ε (6, 7). After the separation of the primates from the other eutherian mammals, η became a pseudogene (7).

After the ancestors of the anthropoids diverged from those of the prosimians ≈55 MYA (8), developmental expression of γ-globin changed from embryonic to fetal in the emerging anthropoids (5, 9). Before this change in developmental expression, the cluster in early primates probably had an arrangement of genes and a pattern of developmental expression resembling that of such strepsirhine prosimians as galago: 5′–ε and γ (embryonic)–ψη (nonexpressed)–δ and β (both fetal and postnatal)–3′ (9, 10). Previous studies (1, 11, 12) indicated the presence in the β-globin cluster of only one γ-globin gene (the expression of which was fetal rather than embryonic) just before the simian or anthropoid radiation leading to platyrrhines (New World monkeys) and catarrhines (humans, apes, and Old World monkeys) and that, after separation of platyrrhines from catarrhines, a 5.5-kb DNA segment encompassing the γ-globin locus was duplicated in the stem catarrhines ancestral to hominoids (humans, gibbons, and other apes) and Old World monkeys. This conclusion regarding when the gene ancestral to $\gamma^1$- and $\gamma^2$-globin tandemly duplicated relied on observations revealing only one γ-globin locus in platyrrhines (1, 12).

In this paper, we demonstrate that spider monkey (a platyrrhine) has two γ-globin genes. Like those of catarrhines, the history of these two γ-globin genes can be traced back to a single γ-globin locus that tandemly duplicated in a common ancestor of platyrrhines and catarrhines. In addition we show that this γ-globin gene duplication resulted from an unequal exchange between two L1 elements flanking the single ancestral γ-globin gene, and we discuss the possible involvement of this duplication in the evolutionary changes that converted embryonic γ-globin genes into fetal γ-globin genes.

## MATERIALS AND METHODS

**Materials.** Enzymes were obtained from Bethesda Research Laboratories, United States Biochemical, and New England Biolabs. Radioactive nucleotide [α-$^{32}$P]dATP was obtained from ICN. X-ray roll film (Kodak XAR-351) and electrophoresis apparatus were obtained from Fotodyne (New Berlin, WI). Wedge spacers (60 cm long, with thickness linearly increasing from 0.2 mm at the top to 0.4 mm at the bottom) for field-gradient gel electrophoresis and acrylamide were obtained from C.B.S. Scientific (Delman, CA).

**Molecular Clones.** Most of the plasmid and phage clones sequenced in this study have been described elsewhere as follows: AgeCh35-19-2 and pAge19-2-H3.4 (in pBR328) from spider monkey (12); HlaCh40-14.5, HlaCh40-12.5; pHla14.5-R8.0, pHla14.5-R2.3, pHla14.5-R1.8, and pHla12.5-K6.5 from gibbon (13); MmuCh30-13.1, MmuCh32-4, MmuCh32-14.7, pMmu13.1-R8.8, pMmu14.7-R2.6, and pMmu14.7-R10.0 from rhesus monkey (14); λGcr15.4 and pGcr15.4-E3.8 from galago (10). One plasmid subclone specifically obtained

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

Abbreviation: MYA, million years ago.

†Present address: Department of Molecular Genetics, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461.

§Present address: Department of Human Genetics, Howard Hughes Medical Institute, University of Michigan Medical Center, Ann Arbor, MI 48109.

for this study, pAge19-2-EB3.1 (in pUC19), included a 3.1-kb *Eco*RI/*Bam*HI fragment of AgeCh35-19-2 containing sequences starting just downstream from the truncated $\gamma^1$-globin locus of spider monkey.

**DNA Sequencing.** Enzymatic sequencing was performed on cloned plasmid or phage DNA, using Sequenase as specified by the vendor (United States Biochemical), and using primers synthesized by the Department of Biochemistry macromolecular synthesis facility at Wayne State University and by the DNA synthesis facility at the Upjohn Company. Primers were designed using the human $\gamma$-globin sequence (11) such that sequencing reactions primed from one strand would overlap adjacent sequencing reactions by 100–300 base pairs (bp). Also, the placement of primers on one strand was staggered with respect to the complementary strand. The library of primers designed from the human sequence was sufficient for obtaining ≈85% of the sequences from the other anthropoid species (gibbon, rhesus monkey, and spider monkey); species-specific primers had to be designed for obtaining some rhesus monkey and spider monkey sequences and most of the galago sequences. Conditions for field-gradient gel electrophoresis to ensure long sequence reads (600–900 bp) have been described (14).

**Evolutionary Analysis.** Nucleotide sequence relationships were determined by parsimony analysis (15) using PAUP (Parsimony Analysis Using Parsimony; David Swofford, Illinois Natural History Survey, Champaign, IL) or by a cluster analysis using the neighbor-joining algorithm (16). In all analyses performed in this work, insertion or deletion events (gaps) were treated as single events, regardless of length, counted at one of the nucleotide positions that they may cover in the sequence alignment. Site-by-site ancestral reconstructions were obtained by hand as described (13).

## RESULTS

**Evidence for Two Linked $\gamma$-Globin Genes in Spider Monkey.** Previous evidence for only a single $\gamma$-globin locus in spider monkey relied on the use of a DNA segment from intron 2 of the cloned spider monkey $\gamma$-globin locus as a probe to spider monkey genomic blots (12). Since several different restriction enzyme digests each yielded only a single band detected with this probe, the investigators concluded that only a single $\gamma$-globin locus existed in the spider monkey genome (12). As part of our present study, we have continued their sequencing effort much further upstream and have found a second, but nonfunctional, $\gamma$-globin gene (Fig. 1). To align the spider monkey sequences with corresponding sequences from catarrhine primates (Fig. 1), a 1.8-kb gap must be placed in the sequence of the upstream locus. This gap delineates a large deletion that occurred in the ancestry of the spider monkey upstream $\gamma$-globin locus and explains the previous failure to find this second upstream $\gamma$-globin locus.

**A Single Duplication Produced the $\gamma^1$- and $\gamma^2$-Globin Genes of Simian Primates.** Having established that spider monkey has two $\gamma$-globin loci, we next addressed the question of whether or not the spider monkey upstream locus was orthologous to the upstream $\gamma^1$-globin locus in catarrhine primates and, similarly, whether or not the spider monkey downstream locus was orthologous to the catarrhine downstream $\gamma^2$-globin locus (Fig. 1). That is, did the $\gamma$-globin duplication precede the divergence of catarrhines (humans, apes, and Old World monkeys) from platyrrhines (New World monkeys, such as the spider monkey), contradicting previous estimates (1, 11, 12), or did the catarrhine $\gamma$-globin loci arise after the catarrhine/platyrrhine divergence from a duplication independent of that producing the two spider monkey loci? To answer this question, we treated as separate sequences the two contiguous, duplicated 5.5-kb segments of related DNA from each of the four simians (human, gibbon, rhesus monkey, and spider monkey), one segment containing

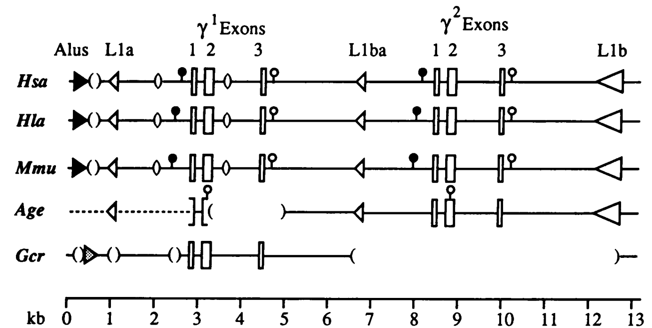
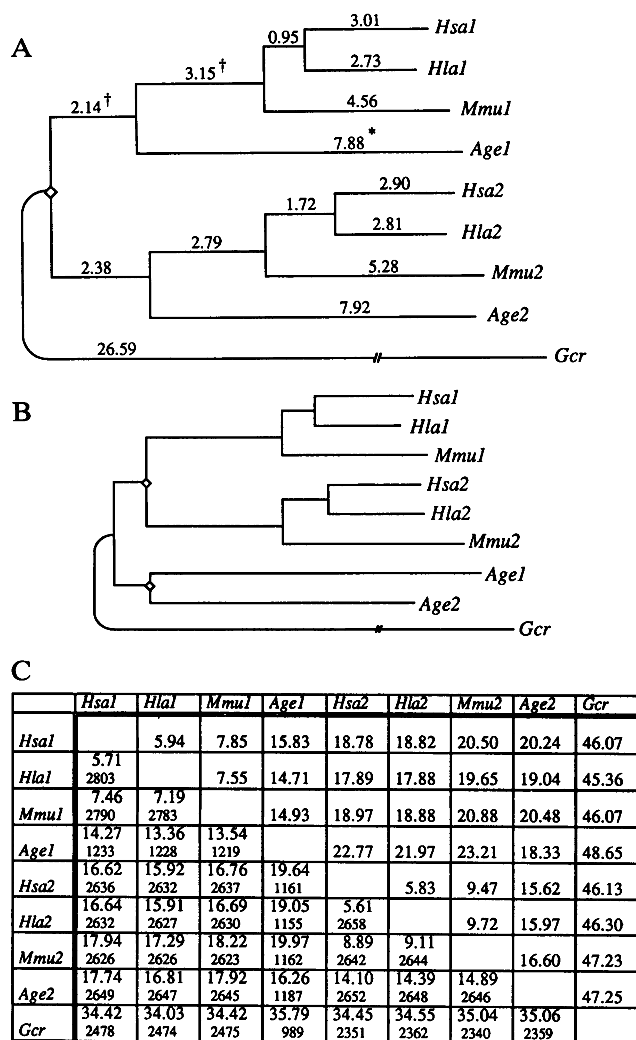



FIG. 1. Schematic of alignment of genomic segments encompassing the duplicated $\gamma$-globin loci of human (*Homo sapiens; Hsa*), gibbon (*Hylobates lar; Hla*), rhesus monkey (*Macaca mulatta; Mmu*), spider monkey (*Ateles geoffroyi; Age*), and the single $\gamma$-globin locus of galago (*Galago crassicaudatus; Gcr*). Parentheses represent end points of the larger alignment gaps. Because galago has only one $\gamma$-globin gene, it has been arbitrarily aligned in this figure with the $\gamma^1$-globin gene of the anthropoids. Open boxes represent $\gamma$-globin exons (the $\psi\gamma^1$-globin gene of spider monkey is truncated within exon 2). Dashed line represents sequences incompletely determined for this paper; subsequent sequencing (K. Hayasaka, D.H.A.F., J.L.S., and M.G., unpublished data) confirms orthology of spider monkey L1a to catarrhine L1a sequences. Arrowheads designate sense-strand orientation of interspersed repetitive elements: open arrowheads, L1 elements; solid arrowheads, Alu elements (the Alu in galago is type II). Only the upstream-most (solid circles) and downstream-most (open circles) boundaries of previously determined (13) converted regions are shown. The sequences described here are from GenBank/EMBL data bases [accession nos. J05174 (*Hla*), X53419 (*Mmu*), X53420 (*Age*), X132286 (*Gcr*), J00179 (previously determined in ref. 11 *Hsa*), and DS5020 (entire annotated alignment)].

the $\gamma^1$-globin locus and the other containing the $\gamma^2$-globin locus. These eight sequences and the single matching sequence from galago (a strepsirhine prosimian and outgroup of Anthropoidea, a group comprising all simians) were then aligned. Prior to analysis, we removed from this alignment a 2.2-kb stretch extending from ≈600 bp 5′ of the initiation codon to the 3′ noncoding region. Gene conversions, the outer boundaries of which are demarcated in Fig. 1, have been shown to have occurred in this region (11, 13, 14) and would have interfered with the reconstruction of the history of the duplication.

Fig. 2*A* shows the most parsimonious tree relating the eight simian sequences from unconverted regions within the duplicated segments and the outgroup galago sequence. This tree depicts the origin of the duplicated $\gamma$-globin genes of both platyrrhines and catarrhines as a single duplication event that preceded the platyrrhine/catarrhine divergence and requires postulating 49 fewer substitutions than the tree depicting the alternative hypothesis of independent $\gamma$-globin duplications in platyrrhines and catarrhines (Fig. 2*B*). Thus, the parsimony criterion strongly supports the conclusion that a single $\gamma$-globin duplication preceded the platyrrhine/catarrhine divergence. This conclusion is also supported by neighbor-joining analysis (16) using the percent divergence values calculated from pairwise comparisons of unconverted sequences within the duplicated segments, corrected (18) for hidden, superimposed substitutions (Fig. 2*C*). The neighbor-joining tree (data not shown) has a branching arrangement that is identical to that of Fig. 2*A* and similar branch lengths. Clearly, the truncated nature of the $\psi\gamma^1$-globin gene of spider monkey (Fig. 1) is due to a large deletion covering much of the original $\gamma^1$-globin gene and not to an independent duplication of only part of the $\gamma$-globin gene.

**Date of the $\gamma$-Globin Duplication.** Although no precise date can be determined for the $\gamma$-globin duplication, a rough estimate can be made by applying local rates of accumulation of nucleotide changes in the unconverted noncoding sequences




| | *Hsa1* | *Hla1* | *Mmu1* | *Age1* | *Hsa2* | *Hla2* | *Mmu2* | *Age2* | *Gcr* |
|---|---|---|---|---|---|---|---|---|---|
| *Hsa1* | | 5.94 | 7.85 | 15.83 | 18.78 | 18.82 | 20.50 | 20.24 | 46.07 |
| *Hla1* | 5.71<br>2803 | | 7.55 | 14.71 | 17.89 | 17.88 | 19.65 | 19.04 | 45.36 |
| *Mmu1* | 7.46<br>2790 | 7.19<br>2783 | | 14.93 | 18.97 | 18.88 | 20.88 | 20.48 | 46.07 |
| *Age1* | 14.27<br>1233 | 13.36<br>1228 | 13.54<br>1219 | | 22.77 | 21.97 | 23.21 | 18.33 | 48.65 |
| *Hsa2* | 16.62<br>2636 | 15.92<br>2632 | 16.76<br>2637 | 19.64<br>1161 | | 5.83 | 9.47 | 15.62 | 46.13 |
| *Hla2* | 16.64<br>2632 | 15.91<br>2627 | 16.69<br>2630 | 19.05<br>1155 | 5.61<br>2658 | | 9.72 | 15.97 | 46.30 |
| *Mmu2* | 17.94<br>2626 | 17.29<br>2626 | 18.22<br>2623 | 19.97<br>1162 | 8.89<br>2642 | 9.11<br>2644 | | 16.60 | 47.23 |
| *Age2* | 17.74<br>2649 | 16.81<br>2647 | 17.92<br>2645 | 16.26<br>1187 | 14.10<br>2652 | 14.39<br>2648 | 14.89<br>2646 | | 47.25 |
| *Gcr* | 34.42<br>2478 | 34.03<br>2474 | 34.42<br>2475 | 35.79<br>989 | 34.45<br>2351 | 34.55<br>2362 | 35.04<br>2340 | 35.06<br>2359 | |

FIG. 2. Phylogenetic trees constructed by using only unconverted sequences within the duplicated genomic segments shown in Fig. 1. Taxa names represent sequences abbreviated as in Fig. 1, where 1 and 2 denote $\gamma^1$- and $\gamma^2$-globin sequences, respectively. (*A*) The most parsimonious tree obtained using the exhaustive search option of PAUP requires postulating a minimum of 1979 total changes. The three next most parsimonious trees (which differed by arrangement of taxa *Hsa1*, *Hla1*, and *Mmu1* with one another, or *Hsa2*, *Hla2*, and *Mmu2* with each other) require eight additional changes each. Branch lengths, shown as percentage divergence (fixed changes per 100 bp), were reconstructed as described elsewhere (17). Lengths were normalized over missing positions in *Age1* (*). The length of the contiguous $\gamma^1$-globin stem anthropoid plus stem catarrhine branch over the large *Age1* deletion was divided proportionately between the separate $\gamma^1$-globin stem anthropoid and stem catarrhine branches (†). (*B*) The alternative hypothesis, that $\gamma$-globin duplications occurred independently in catarrhine (*Hsa, Hla, Mmu*) and platyrrhine (*Age*) lineages, requires a minimum of 2028 changes and is only the 52nd tree encountered on ordering all possible trees according to increasing minimum length. (*C*) Pairwise comparisons of unconverted sequences within the duplicated segments, showing values (% divergence) corrected (18) for hidden, superimposed substitutions above, and uncorrected values (over the number of positions compared) below the diagonal.

within the duplicated segments. Using 25 MYA for the catarrhine radiation date and 37.5 ± 2.5 MYA for the anthropoid radiation date (when the catarrhine and platyrrhine lineages began to diverge) (17, 19) and branch lengths from the most parsimonious tree (Fig. 2*A*: $\gamma^1$-globin, 3.15%; $\gamma^2$-globin, 2.79%) and from the neighbor-joining tree (data not shown; $\gamma^1$-globin, 2.82%; $\gamma^2$-globin, 3.55%), a rate of $2.6 \pm 0.8 \times 10^{-7}$ %/yr is obtained for the $\gamma^1$- and $\gamma^2$-globin catarrhine stems. The mean number of changes accumulating on the anthropoid stems after the $\gamma$-globin duplication is 2.30% ± 0.24% (parsimony tree, Fig. 2*A*: $\gamma^1$-globin, 2.14%; $\gamma^2$-globin, 2.38%; neighbor-joining tree, data not shown: $\gamma^1$-globin, 2.61%; $\gamma^2$-globin, 2.08%). Applying the rate calculated for the catarrhine stems, the date at which the duplication occurred is estimated to be 10.0 ± 4.0 million years before the anthropoid radiation, or about 47.5 ± 6.5 MYA.

**Mechanism of the Duplication.** In the simian sequences (Fig. 1), at each boundary of the contiguous duplicated DNA segments containing the two $\gamma$-globin genes is a highly truncated member of the L1 family of long interspersed repetitive elements (11, 26). Orthologous L1 elements are absent in galago (Fig. 1) and in other outgroup mammals such as rabbit (23). Thus, we propose that early in the lineage to the ancestor of simian primates, after diverging from the strepsirhine prosimians (e.g., galago), but before the $\gamma$-globin duplication, the ancestral L1a element inserted upstream of the single $\gamma$-globin gene while the ancestral L1b element independently inserted downstream of the gene. In one model (Fig. 3*A*), the $\gamma$-globin duplication is proposed to have originated as one of the products of an unequal homologous crossover (11, 26) between mispaired L1a and L1b elements. The reciprocal product, a chromosome containing a deletion of the $\gamma$-globin locus (data not shown), was presumably not fixed. As a result of this unequal exchange, the ancestral L1ba element would have comprised 5′ sequences of the 3′ element, L1b, and 3′ sequences of the 5′ element, L1a (Fig. 3*A*). Otherwise (e.g., if L1ba had inserted independently), it would be very unlikely for L1ba to have had exactly this type of hybrid structure.

That the $\gamma$-globin duplication arose from such an unequal exchange is supported by comparing the aligned L1a, L1ba, and L1b sequences. In Fig. 3*B*, a portion of the simian L1 sequences (actually sequences from the L1 antisense strand, consistent with their schematic orientation in Fig. 1) are aligned with one another and with a human L1 consensus sequence (24), with which they are clearly related. When the ancestral anthropoid L1a, L1ba, and L1b elements and their flanking sequences are reconstructed (Fig. 3*B*), it is clear that the prediction of the unequal exchange model is met. That is, the ancestral L1ba element appears to be a hybrid element consisting of sequences from the 3′ part of L1a (downstream from alignment position 43; Fig. 3*B*) and sequences from the 5′ part of L1b (upstream from position 42), supporting an unequal exchange as the origin of the duplication with breakpoints between positions 42 and 43 in the preduplication L1a and L1b elements. Phylogenetic scanning (20) also supports the unequal exchange model for the origin of the $\gamma$-globin duplication and unequivocally maps the exchange breakpoint (Fig. 3*C*, X) between positions 42 and 44.

## DISCUSSION

About 100–80 MYA tandem duplications of the embryonic proto-$\varepsilon$-globin locus led to $\varepsilon$, $\gamma$, and $\eta$ loci (6, 7). While $\varepsilon$-globin genes continued to function as indispensible, embryonically expressed genes, $\gamma$ and $\eta$ apparently served as extra or redundant $\varepsilon$-globin genes and were dispensible [$\gamma$-globin was deleted in artiodactyls and $\eta$-globin became a pseudogene in primates and was deleted in rodents and lagomorphs (7)]. The $\gamma$-globin genes have continued to function as embryonic genes in rodents, lagomorphs, and prosimian primates, although considerable amino acid sequence divergence has accumulated in the $\varepsilon$- and $\gamma$-globin chains encoded by these embryonically expressed $\varepsilon$- and $\gamma$-globin genes (9, 10). Although much of this divergence may be selectively neutral, some of it may have been adaptive, giving the $\varepsilon$- and $\gamma$-globin genes somewhat different embryonic niches, implying that embryonic $\gamma$-globin genes may no longer be dispensible in animals such as rabbit and galago. However, $\gamma$-globin genes in their

A

Unequal exchange in an Anthropoid ancestor

B

```
Hsa L1a   ACAAGTGCAATATGTGCAGCTTTGTTGCGCAGGTCAACATGTATCTTTCTGGTCTTTT
Hla L1a   ACAAGTGCAATATGTGCAGCTTTGCTGCGCAGGTCAACATGTATCTTTCTGGTCTTTT
Mmu L1a   ACAAGTGCAATATGTGCAGCTTTGCTGCATGGGTCAACATGAATCTTTCTGGTCTTTT
Hsa L1ba  ACATGTGCAGGATGTGCAGGCTTGTTACATGGGTAAATGTGTGTCTTTCTGGCCTTTT
Hla L1ba  ACATGTGCAGGATATGCAGGCTTGTTACATGGGTAAATGTGTATCCTTCTGGCCTTTT
Mmu L1ba  ACATGTGCAGGATGTGCAGGCTTGTTACATGGGCAAATGTGTACCTTTCTGGCCTTTT
Age L1ba  ACATGTGCAAGATATACAGGCCTTTTACATGGGTAAATGTGTATCTTTCTGGCCCTTT
Hsa L1b   ACACGTACAGGATATGCAGGTTTGTTTTATAGGTAAAAGTGTGCCATGGT********
Hla L1b   ACATGTACAGGATATGCAGATTTGTTTTATAGGTAGAAGTGTGCCATGGT********
Mmu L1b   ACATGTGCAGGATATGCAGGTTTGTTTTATAGGTA*ATGTGTGCCATGGT********
Age L1b   ATGTGTGCTGGATATGCAAGCTTGTTTTACAGGTAAAAGTGTGCCATGGT********

Ant L1a   ACAAGTGCAATATGTGCAGCTTTGCTGCGCAGGTCAACATGTATCTTTCTGGTCTTTT
Ant L1ba  ACATGTGCAGGATATGCAGGCTTGTTACATGGGTAAATGTGTATCTTTCTGGCCTTTT
Ant L1b   ACATGTGCAGGATATGCAGGTTTGTTTTATAGGTAAAAGTGTGCCATGGT********
                    unequal exchange breakpoint
cons.     ACATGTGCACAACGTGCAGGTTTGTTACATATGTATACATGTGCCATGTTGGTGTGCT
position  1--------+---------+---------+---------+---------+------58
```

C

FIG. 3. (A) Unequal exchange between mispaired ancestral L1a and L1b elements in the origin of the γ-globin gene duplication. (B) Partial sequence alignment of primate L1 elements (abbreviated and oriented as in Fig. 1) to a human consensus L1 sequence (cons.) (24). Asterisks denote gaps introduced to maximize alignment. Identities between the reconstructed anthropoid ancestral L1ba element (Ant L1ba) and either the ancestral L1a or the ancestral L1b element are indicated with vertical lines. Ancestral sequences were reconstructed as described (13) and actually represent a subset of equally likely reconstructions. Interestingly, these L1s are most closely related to a human L1 type J consensus sequence postulated to be the most ancient L1 group (25). (C) The "phylogenetic scanning" (20) algorithm designed to test the model (A) determines in 20-base intervals whether L1b and L1ba are more closely related than either is to L1a (the 5′ tree; *Lower Left*), or whether L1a and L1ba are more closely related than either is to L1b (3′ tree; *Lower Right*). The spider monkey L1a sequence (*Age* L1a) was unavailable at the time of the analysis; dashed line represents the putative position of the *Age* L1a branch. Because no appropriate L1 outgroup was available, the degree to which ancestral L1 lineages are related was measured by relative lengths of branches α (open bars) and β (solid bars), such that the tree root was placed on the longer of these two branches. The abscissa of the plot denotes the 3′-most position (in the L1 alignment; B) of the 20-base window within which unambiguous lengths of α (○) or β (▲) were evaluated. Wherever α > β, L1ba is more closely related to L1b than to L1a (the 5′ tree); β > α indicates that L1ba is more closely related to L1a than to L1b (the 3′ tree). The unequal exchange breakpoint is mapped by this method to a point (X) between positions 42 and 44 (B). When changes with ambiguous branch assignment are included in the analysis, even stronger support for this unequal exchange is obtained due to the greater differences between branch lengths α and β.

role as extra embryonic genes were apparently redundant in the early primates and, in the ancestral anthropoid lineage about 50–35 MYA, were recruited to be fetally expressed genes (9). The time that this recruitment occurred is intriguingly close to the time that $\gamma^1$- and $\gamma^2$-globin genes arose from the tandem duplication mediated by L1 elements flanking the ancestral single γ-globin locus.

Our phylogenetic reconstruction shows that during this time, 14 nonsynonymous substitutions, as compared to only 6 synonymous substitutions, accumulated in the stem anthropoid γ-globin genes (Fig. 4). Some of the resulting amino acid changes restructured the encoded γ-globin chains for the functional requirements of fetal hemoglobin (5, 9). In addition, there were nucleotide changes in the 5′ promoter region of the stem anthropoid γ-globin genes that are implicated in the regulatory changes that delayed expression of these genes from embryonic to fetal life (9, 21). We know that concerted evolution via frequent gene conversions occurs not only in the coding sequences of the paired $\gamma^1$- and $\gamma^2$-globin genes, but also in their 5′ regulatory sequences (11, 13, 14). Thus, if the duplication resulting in $\gamma^1$- and $\gamma^2$-globin genes had occurred before they were recruited as fetal genes, concerted evolution plus selection for favorable mutations would have spread the regulatory and coding mutations responsible for fetal expression and function to both genes.

We can conjecture but not prove that duplication of the γ-globin genes facilitated (and thus accelerated) the transition from embryonic to fetal function by doubling the number of extra, unneeded embryonic genes that were freer to collect the changes required for fetal expression and function. This hypothesis rests on the observation that changes can accumulate in either gene duplicate not only by mutation, but also by intergenic (as well as interallelic) recombination (22, 27). Whether or not duplication of the γ-globin genes played such a role, two genes are probably not needed for continued occupation of this fetal niche. In catarrhines, the level of $\gamma^2$-globin expression in fetal life is 3 times lower than that of $\gamma^1$-globin (28) and the ratio of nonsynonymous to synonymous changes is generally higher in catarrhine $\gamma^2$- than in $\gamma^1$-globin lineages (Fig. 4). This pattern suggests relaxed selection on catarrhine $\gamma^2$-globin genes and the possibility that they may eventually become nonfunctional. In platyrrhines, the lineage of the expressed spider monkey $\gamma^2$-globin gene displays a ratio of nonsynonymous to synonymous substitutions that is the same as, if not higher than, that of the nonexpressed spider monkey $\psi\gamma^1$-globin lineage (Fig. 4). Perhaps in platyrrhines, $\gamma^2$-globin was initially relatively unconstrained while $\gamma^1$-globin was the selectively constrained gene until a large deletion inactivating $\gamma^1$-globin was fixed in the spider monkey lineage, leaving $\gamma^2$-globin gene as a backup copy.

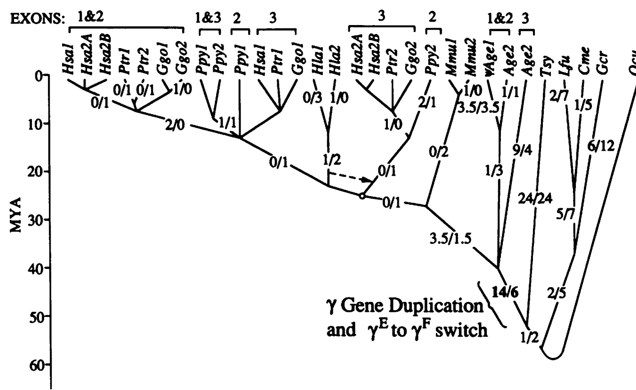



FIG. 4. Ratios of amino acid changing to silent substitutions during evolution of the coding regions of primate γ-globin genes. Besides γ-globin sequences from the species shown in Fig. 1, other sequences (13) included are from the human $\gamma^2$-globin B allele (*Hsa2B*), chimpanzee (*Ptr*), gorilla (*Ggo*), orangutan (*Ppy*), tarsier (*Tsy*), brown lemur (*Lfu*), dwarf lemur (*Cme*), and rabbit (*Ocu*). Branching patterns and approximate nodal dates depicted are those determined previously (5, 13). The reconstructed branch lengths represent the number of nonsynonymous (numerator) and synonymous (denominator) substitutions postulated to have occurred between ancestral and descendant sequences for their total number of aligned positions (3, 5). These branch lengths are only presented where they are nonzero. The complex branching arrangement reflects the varying extents of gene conversion between $\gamma^1$- and $\gamma^2$-globin genes in particular species lineages (5, 13). Gorilla $\gamma^1$-globin (*Ggo1*) and $\gamma^2$-globin (*Ggo2*) share a conversion up to but not including the last codon of exon 2; this codon is thus included in the exon 3 taxon. Regions of *Age2* not shared with *Age1* are grouped under the exon 3 taxon. *Age1* and *Age2* share a gene conversion nearly the entire length of the *Age1* exon sequences. Note the node (○) designating a conversion in a hominoid ancestor (13). Dashed arrow shows alternative placement of the gibbon γ-globin lineage.

This multifaceted history of the duplicated anthropoid γ-globin genes, marked both by genes that serve an important new functional role and by those that have become nonfunctional pseudogenes, exemplifies the fates of duplicated genes. It also provides a model for gene duplication involving interspersed repetitive elements at the site of unequal exchange. The large number of examples documenting unequal exchanges involving such elements (29–31) lends credibility to the suggestion that unequal exchanges between such elements may contribute significantly to the formation of duplications (30, 32). It has been suggested (33) that natural selection acts through the long-term advantage of functional diversity to maintain the ability granted by unequal exchange to cause gene duplications. It has also been observed that unequal exchange has a role in controlling copy number and preventing the loss of interspersed repetitive elements (34). By being numerous, interspersed, and homogeneous (similar in sequence), interspersed repetitive elements such as L1s can be nucleation sites for these unequal exchanges. We speculate that, if the short-term genetic load associated with such elements (e.g., insertions into essential genes) is not too high, their presence may actually provide genomes that harbor them a long-term advantage by their ability to mediate gene duplications.

**Note Added in Proof.** A recently discovered conversion between *Age1* and *Age2* 3′ of the coding region reduces those branch lengths somewhat but alters no conclusions.

We thank Kenji Hayasaka for sequence confirmation that *Age* L1a exists, and W. Bailey, S. Baird, J. Crow, G. Dover, S. Emmons, T. Grudzien, S. Hawley, B. Koop, M. Miyamoto, N. Moncrief, T. Ohta, P. Rogan, M. Stanhope, L. Straussbaugh, and two anonymous reviewers for their many helpful comments and suggestions. This research was supported by the National Institutes of Health and the National Science Foundation.

1. Barrie, P. A., Jeffreys, A. J. & Scott, A. F. (1981) *J. Mol. Biol.* **149**, 319–336.
2. Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C. & Proudfoot, N. J. (1980) *Cell* **21**, 653–668.
3. Czelusniak, J., Goodman, M., Hewett-Emmett, D., Weiss, M. L., Venta, P. J. & Tashian, R. E. (1982) *Nature (London)* **298**, 297–300.
4. Hardison, R. C. (1984) *Mol. Biol. Evol.* **1**, 390–410.
5. Goodman, M., Czelusniak, J., Koop, B. F., Tagle, D. A. & Slightom, J. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52**, 875–900.
6. Koop, B. F. & Goodman, M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 3893–3897.
7. Goodman, M., Koop, B. F., Czelusniak, J., Weiss, M. L. & Slightom, J. L. (1984) *J. Mol. Biol.* **180**, 803–823.
8. Koop, B. F., Tagle, D. A., Goodman, M. & Slightom, J. L. (1989) *Mol. Biol. Evol.* **6**, 580–612.
9. Tagle, D. A., Koop, B. F., Goodman, M., Slightom, J. L., Hess, D. L. & Jones, R. T. (1988) *J. Mol. Biol.* **203**, 439–455.
10. Tagle, D. A. (1990) Ph.D. thesis (Wayne State Univ. School of Medicine, Detroit).
11. Shen, S.-H., Slightom, J. L. & Smithies, O. (1981) *Cell* **26**, 191–203.
12. Giebel, L. B., van Santen, V. L., Slightom, J. L. & Spritz, R. A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 6985–6989.
13. Fitch, D. H. A., Mainone, C., Goodman, M. & Slightom, J. L. (1990) *J. Biol. Chem.* **265**, 781–793.
14. Slightom, J. L., Koop, B. F., Xu, P. & Goodman, M. (1988) *J. Biol. Chem.* **263**, 12427–12438.
15. Fitch, W. M. (1971) *Syst. Zool.* **20**, 406–416.
16. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
17. Bailey, W. J., Fitch, D. H. A., Tagle, D. A., Czelusniak, J., Slightom, J. L. & Goodman, M. (1991) *Mol. Biol. Evol.* **8**, 155–184.
18. Jukes, T. H. & Cantor, C. R. (1969) in *Mammalian Protein Metabolism*, ed. Munro, H. N. (Academic, New York), pp. 21–123.
19. Gingerich, P. D. (1984) *Yearb. Phys. Anthropol.* **27**, 57–72.
20. Fitch, D. H. A. & Goodman, M. (1991) *CABIOS* **7**, 207–215.
21. Gumucio, D. L., Blanchard-McQuate, K. L., Helstedt-Williamson, H., Tagle, D. A., Gray, T. A., Tarle, S. A., Grogowski, L., Goodman, M., Slightom, J. L. & Collins, F. S. (1991) in *Proceedings of the 7th Conference on Hemoglobin Switching*, eds. Stomatoyannopoulos, G. & Nienhuis, A. W. (John Hopkins Univ. Press, Baltimore), in press.
22. Dover, G. (1987) *J. Mol. Evol.* **26**, 47–58.
23. Margot, J. B., Demers, G. W. & Hardison, R. C. (1989) *J. Mol. Biol.* **205**, 15–40.
24. Scott, A. F., Schmeckpeper, B. J., Abdelrazik, M., Comey, C. T., O'Hara, B., Rossiter, J. P., Cooley, T., Heath, P., Smith, K. D. & Margolet, L. (1987) *Genomics* **1**, 113–125.
25. Jurka, J. (1989) *J. Mol. Evol.* **29**, 496–503.
26. Maeda, N. & Smithies, O. (1986) *Annu. Rev. Genet.* **20**, 81–108.
27. Maizels, N. (1989) *Trends Genet.* **5**, 4–8.
28. Bunn, H. F. & Forget, B. G. (1986) in *Hemoglobin: Molecular, Genetic and Clinical Aspects* (Saunders, Philadelphia), p. 68.
29. Davis, P. S., Shen, M. W. & Judd, B. H. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 174–178.
30. Goldberg, M. L., Sheen, J.-Y., Gehring, W. J. & Green, M. M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5017–5021.
31. Lehrman, M. A., Goldstein, J. L., Russell, D. W. & Brown, M. S. (1987) *Cell* **48**, 827–835.
32. Barsh, G. S., Seeburg, P. H. & Gelinas, R. E. (1983) *Nucleic Acids Res.* **11**, 3939–3958.
33. Ohta, T. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 3509–3512.
34. Langley, C. H., Montgomery, E., Hudson, R., Kaplan, N. & Charlesworth, B. (1988) *Genet. Res.* **52**, 223–235.