Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 3893–3897, June 1988
Evolution

Evolutionary and developmental aspects of two hemoglobin β-chain genes (ε^M and β^M) of opossum

BEN F. KOOP*† AND MORRIS GOODMAN‡

*Department of Molecular Biology and Genetics and ‡Department of Anatomy and Cell Biology, Wayne State University, Detroit, MI 48201

Communicated by Roy J. Britten, November 30, 1987 (received for review August 26, 1987)

ABSTRACT **A series of gene duplications that began in a stem species of Mammalia and led to five developmentally regulated hemoglobin β-chain loci (ε, γ, η, δ, and β) in a common ancestor of eutherian orders Artiodactyla, Rodentia, Lagomorpha, and Primates had important consequences in mammalian evolution. Findings reported here indicate that two progenitors of the five linked genes existed by the time of the eutherian (placental mammal)–metatherian (marsupial mammal) split and that these two genes were already differentiated with respect to their promoter regions and developmental expression. Southern blot and sequence analyses of the hemoglobin β-chain genes of the opossum (*Didelphis virginiana*) revealed only two genes, one with coding and promoter sequences similar to eutherian prenatally expressed ε, γ, and η genes and the other coding for adult opossum hemoglobin β-chains and having eutherian adult β-type promoters. The most parsimonious arrangement of >80 β-globin exon sequences depicts the opossum embryonic-type gene as orthologously related to eutherian ε, γ, and η genes and the opossum adult-type gene as orthologously related to δ and β genes. These data further indicate that after the initial β duplication in the stem of Mammalia, the locus that became developmentally delayed in its expression evolved at a faster rate than the locus that became embryonically expressed.**

Previous studies have provided evidence that a series of tandem gene duplications during formative periods of mammalian evolution led to a cluster of five developmentally regulated hemoglobin β-chain loci (5′ ε–γ–η–δ–β 3′) in which ε, γ, and η were embryonically expressed genes and δ and β were adult-expressed genes (1–5). The first of the duplications occurred in the range of 155–200 million years ago (1, 6–8) and produced two gene lines, of which one was the progenitor of ε, γ, and η loci and the other was the progenitor of δ and β loci (2–4). The later duplications that separated ε, γ, and η from each other and in parallel δ and β were estimated to have occurred in the range of 90–140 million years ago (1, 6–8).

The proposed range of dates for the later duplications coincides with the geological period (100–135 million years ago) in which the stem-marsupials and stem-eutherians are thought to have diverged from their last common ancestor (9–11). Clearly, the β-chain-related hemoglobin genes of marsupials need to be examined and compared to those of eutherians to determine if more than two progenitors of ε, γ, η, δ, and β loci existed at the time of the metatherian–eutherian split and whether the last common ancestor of marsupials and placental mammals already had β-chain-related hemoglobin genes that were differentially expressed in embryonic and adult life.

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.

MATERIALS AND METHODS

High molecular weight DNA from *Didelphis virginiana* (female, from Southfield, MI) liver tissue was used to make a Charon 35 λ phage recombinant library (12). The genomic library (≈1 × 10^6 plaque-forming units) was screened by using human ε, human γ, gorilla γ, human β, and goat η gene probes (12, 13). Hybridization conditions were 58°C, 1 M NaCl, and 1% $NaDodSO_4$. Washing conditions were 58°C, 0.15 M NaCl/15 mM sodium citrate, and 0.5% $NaDodSO_4$. Positive clones were isolated and mapped with *Bam*HI, *Eco*RI, and *Hin*dIII, and gene regions were localized by Southern blot analysis (13). Two gene regions were isolated (see Fig. 1). To confirm the integrity of the cloned regions and to determine if additional β-globin genes could be detected, *Bam*HI, *Bgl* II, *Eco*RI, and *Hin*dIII restriction digests of opossum genomic DNA were electrophoresed alongside a non-human primate control species (*Galago crassicaudatus*, *Eco*RI digest), blotted, and hybridized to exon 1 and 2 probes from human ε, human γ, human β, and the two opossum genes [under conditions as stated above except that hybridizations were performed at 40°C (in the case of human β, 36°C)]. Gene-containing *Eco*RI fragments from Dvi Ch35-14.8 and Dvi Ch35-14.4 (Fig. 1) were subcloned into pUC8 plasmids and grown in JM83 hosts. Nucleotide sequences were obtained by using chemical cleavage methods (14) as described by Slightom *et al.* (12). Phylogenetic reconstruction procedures are described by Goodman *et al.* (4). Synonymous and nonsynonymous gene divergence values were calculated according to Nei and Gojobori (15).

RESULTS AND DISCUSSION

Genomic and λ Clone Maps of Two Opossum β-Globin Gene Regions. Recombinant clones Dvi Ch35-14.8, Dvi Ch35.14.4, and Dvi Ch35-18.0 were isolated and mapped, and globin-containing fragments were identified by Southern blot analysis (Fig. 1). To confirm the integrity of cloned regions and to determine if additional β-globin genes could be identified, genomic DNA of opossum and a non-human control species (*G. crassicaudatus*) were probed with exon sequences from human ε, γ, and β as well as the two opossum genes. These probes detected only those fragments also found in cloned sequences (under hybridization conditions that detected all functional β-globin genes within the control DNA).

We found no evidence in the genomic library or in opossum genomic blots for the presence of more than two opossum β-globin genes, although this does not exclude the possibility of pseudogenes or anciently separated lineages being present. Our results suggest only that if more than two β-globin genes occur in the opossum genome, they are more divergent than the most divergent functional β-globin genes found in eutherians and more divergent than the two isolated opossum β-globin genes.

†To whom reprint requests should be addressed.

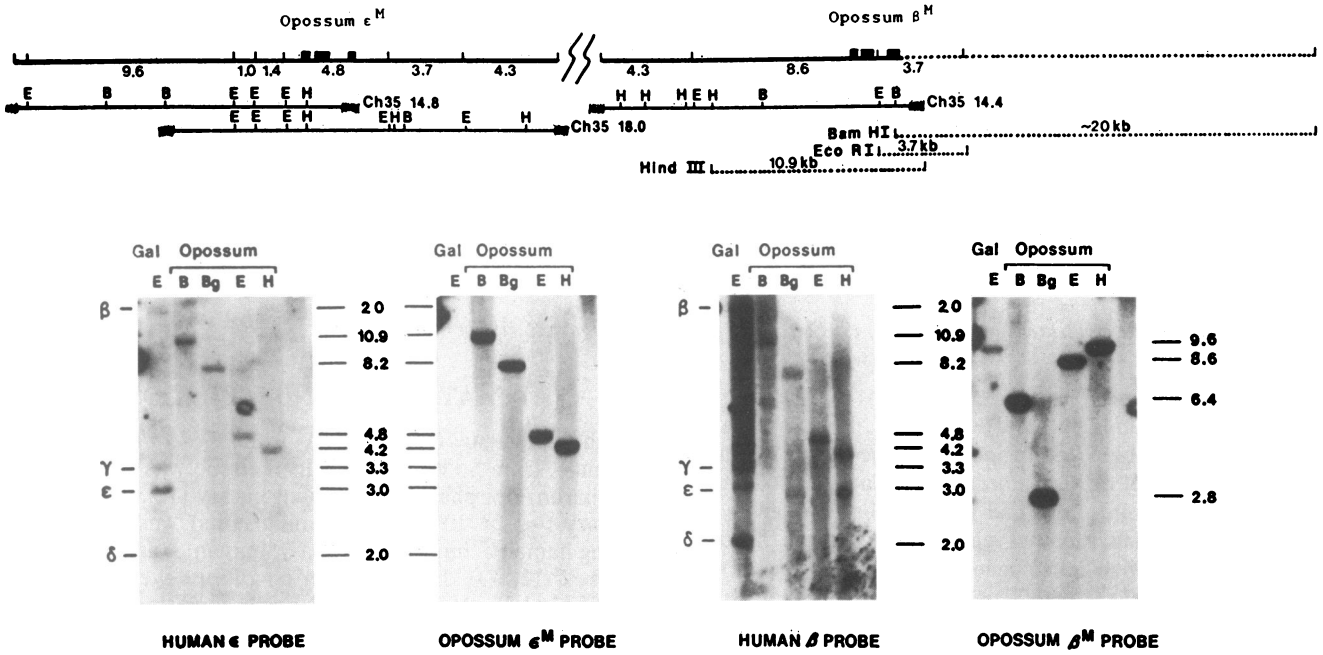

FIG. 1. Genomic and λ clone maps of opossum ε^M- and β^M-globin gene regions. The genomic map (upper line) was established by Southern blot analysis of recombinant λ clones and total genomic DNA restriction digests. The genomic map was extended (dotted line) by probing total genomic DNA, digested with the above restriction enzymes, with a 700-base *Bam*HI fragment isolated from the most 3′ end of the Dvi Ch35-14.4 insert. *Bam*HI (B), *Bgl* II (Bg), *Eco*RI (E), and *Hin*dIII (H) restriction digests of opossum and galago (Gal) genomic DNA were hybridized to exon 1 and 2 probes from human ε, human γ (not shown), human β, and both opossum genes (under conditions stated in the text). Human ε, γ, and β probes not only clearly hybridized to their orthologous counterpart present in galago but also to paralogous genes present in galago [ε, 3.0 kilobases (kb); γ, 3.3 kb; δ, 2.0 kb; β, 20 kb]. In opossum digests, however, human ε (and γ) probes hybridized only to single fragments corresponding to fragment sizes found in recombinant clone Dvi Ch35-18.0 (fragment sizes are given in kb); human β probes failed to unambiguously hybridize with any opossum DNA fragments at comparable stringencies (data not shown); however, at reduced stringencies human β probes hybridized to fragments with sizes found in clones Dvi Ch35-18.0 and Dvi Ch35-14.4. Opossum ε^M and β^M probes hybridized only to single fragments with sizes found in the clones from which the probes were obtained. The opossum probes hybridized to the same opossum DNA fragments identified by human ε and γ probes. Although opossum β^M hybridizes to a galago *Eco*RI fragment (10.6 kb), that fragment is not recognizable as a β-globin gene and is considered an artifact attributable to reduced stringencies of hybridization.

The nucleotide sequences of the two opossum gene regions[§] are presented in Fig. 2. For reasons that will become apparent (next section) we have labeled them β^M and ε^M. Only in the coding and 5′ regulatory regions could we find sequence alignments indicative of homology either between opossum ε^M and β^M genes themselves or between opossum and eutherian genes. Regions of homology show that both opossum genes have the same promoter and intron–exon organization as other vertebrate β-globin genes.

Opossum gene β^M translates into the known adult opossum hemoglobin β-chain amino acid sequence (9) (except at position 80 of the amino acid sequence where asparagine occurs instead of aspartic acid). Opossum gene ε^M as judged by its sequence is functional, although a second β-globin has not been identified in adult or juvenile opossums (9, 19).

Sequence Comparisons. The coding region of opossum adult β^M differs at nonsynonymous positions from human ε, human γ (human η is a pseudogene and therefore is not included in these comparisons), human β, and a partial sequence of an Australian marsupial cat (*Dasyurus viverrinus*) adult β (20) by 20%, 24%, 23%, and 12%, respectively (Table 1). With the exception of the divergence between Australian and North American marsupials, these values are greater than those found between different human β-globin genes (ε–γ, ε–β, and γ–β diverge by 10%, 16%, and 18%, respectively) and are very similar to values between chicken and eutherian β-globins (Table 1). Though the opossum adult β-globin gene is functional, it is clearly very divergent from eutherian β-globin genes. For example, it diverges more from *Xenopus* β than do any of the human genes (Table 1).

The second opossum β-globin gene (ε^M) appears very highly conserved and most closely resembles the ε- and γ-globins of humans. Its coding region differs at nonsynonymous positions from human ε, human γ, human β, and the adult opossum β by 13%, 13%, 16%, and 22%, respectively. That the human ε and γ genes are more similar to each other than either is to opossum β^M and ε^M indicates that the eutherian ε and γ gene lineages may have split after the separation of eutherians and metatherians.

The 5′ promoter region of opossum β^M resembles that of eutherian adult-type β-globin promoters in having two characteristic types of CACA boxes (2, 5, 18, 21) located 9 bp and 24 bp upstream of the CAAT box (Fig. 3). It most closely resembles the promoter region found in mouse β-major (Fig. 3). The 5′ promoter region of the non-adult opossum ε^M locus, however, closely resembles the eutherian ε-, γ-, and η-globin promoter regions in the vicinity of the CAAT and single CACA box regions (Fig. 3). The non-adult opossum CACA box has the distinctive sequence CTCCACCCC found in the promoter region of the human ε-globin gene and, with minor variations, in the promoter regions of other eutherian ε, γ, and η genes. Like human and rabbit γ genes (1), opossum ε^M has two CAAT promoter regions, with an identical sequence (TGACCAAT) spanning the 3′ CAAT box in opossum and eutherians. The 5′ CAAT boxes, however, appear to have arisen independently in eutherians and opossum (Fig. 3). With the exception of fetally expressed simian primate γ genes, functional eutherian ε, γ, and η genes are expressed during embryonic development (1, 3, 5, 7, 22, 23).

[§]The sequences reported here are being deposited in the EMBL/GenBank Genetic Sequence Database (Intelligenetics, Mountain View, CA) (accession no. J03642 for ε^M and accession no. J03643 for β^M).

FIG. 2. Nucleotide and deduced amino acid sequence of opossum ε^M (*Upper*) and β^M (*Lower*) genes. Both opossum genes are divided into three exons separated by two introns. The immediate 5′ flanking regions of both sequences contain CACCC and CCAAT types of promoter elements and an RNA polymerase II binding site (ATA). Both sequences share the same initiation (INI) and terminator (TER) sequences as well as a 3′ poly(A) addition signal (AATAAA). In the extended 5′ region of β^M, located 407 base pairs (bp) 5′ of the initiation codon, is a purine/pyrimidine dinucleotide repeat similar to that found about 400 bp 5′ of *Xenopus* adult β1 (16) and that found 578 bp 5′ of human adult β (17). The length of intron 1 in both opossum β genes (109 bp and 113 bp in ε^M and β^M, respectively) is similar to that of other vertebrate β genes, but intron 2 is substantially larger. Intron 2 lengths in ε^M and β^M are 1857 and 1465 bp, respectively, whereas the largest vertebrate β-globin intron 2 found in previous studies was in goat ε^{II} [1040 bp, (18)].

Although a protein chain encoded by opossum ε^M has not been found in juvenile or adult opossums, the non-adult opossum gene is conserved to a greater extent than the gene expressed in adult opossums. Inasmuch as very similar eutherian ε-, γ-, and η-globin genes are expressed during prenatal development, we deduce that the non-adult opossum gene, like its eutherian counterparts, encodes an embryonic hemoglobin chain. Thus, we designated the two opossum genes as β^M (marsupial adult-type) and ε^M (marsupial embryonic-type).

Table 1. Nonsynonymous divergence values (percent) over pairs of globin gene coding regions

Human ε	—							
Human γ	10	—						
Opossum ε^M	13	13	—					
Opossum β^M	20	24	22	—				
Marsupial cat β*	16	18	17	12	—			
Human β	16	18	16	23	18	—		
Chicken β	19	23	21	24	21	23	—	
Xenopus β	33	35	37	43	48	38	40	—

*Partial sequences.

The most parsimonious genealogical arrangement for >80 β-globin genes (representative lineages shown in Fig. 4) joins opossum β^M to the root of a branch leading to eutherian β genes and joins opossum ε^M to the root of a separate branch leading to eutherian ε, γ, and η genes. The two paralogous branches (represented by ε^M and β^M in opossum) trace back to a single gene duplication that occurred after the ancestral

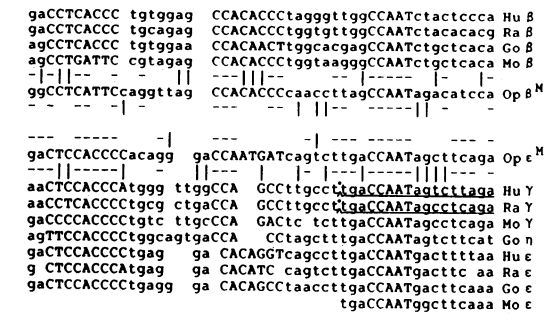

FIG. 3. Comparison of 5′ flanking promoter sequences of β-related globin genes. Highly conserved sequences that have been implicated in gene regulation (CACA; CCAAT boxes) and their orthologous counterparts in other sequences are given in uppercase letters. Underlined human and rabbit γ sequences are parts of 27-base imperfect repeats involving the CCAAT box: in each case, the 3′ duplicate, which better aligns with other sequences, is presented. Vertical bars above, or below, opossum sequences indicate nucleotides in the opossum that are also found in at least 50% of the eutherian sequences above it (adult β genes) or below it (ε, γ, and η genes). Horizontal bars indicate positions where at least 50% of the eutherian sequences above *and* below the opossum sequences have the same nucleotide. The opossum β^M sequence clearly aligns best with eutherian adult β 5′ flanking regions, and the opossum ε^M sequence aligns best with eutherian ε, γ, and η 5′ flanking regions. Hu, Ra, Go, Mo, and Op, human, rabbit, goat, mouse, and opossum, respectively.

separation of Aves and Mammalia but before the separation of Metatheria and Eutheria. Alternative branching arrangements for the origins of opossum ε^M and β^M require at least three extra genic events.

In the most parsimonious tree (Fig. 4), after the orthologous separation of opossum ε^M from antecedents of eutherian ε, γ, and η loci and before the separation of artiodactyl, primate, lagomorph, and rodent lineages, two duplications first separate out ε and then η and γ loci. Emergence of the eutherian embryonic-type loci is compatible with their physical linkage arrangement, which is 5′ ε–γ–η 3′. Alternative arrangements require at least four extra nucleotide substitutions. Also in the most parsimonious tree, Artiodactyla separates first from other eutherian orders, Primates second, and Lagomorpha and Rodentia last. This pattern parallels superordinal relationships proposed in several recent molecular and morphological studies (26–28). Alternative arrangements require at least nine extra substitutions. At least nine extra substitutions are also required to prevent opossum β^M and the Australian marsupial cat from forming a monophyletic group.

The results presented in Fig. 4 are compatible with (*i*) the number and 5′ → 3′ linkage order of eutherian β globin genes, (*ii*) differences in divergence rates determined from pairwise comparisons, (*iii*) types and locations of putative promoter sequences, and (*iv*) patterns of developmental expression of ε and β-like globin genes in mammals. The separate lines of evidence by themselves do not conclusively point to any single chain of events but together are consistent with the hypothesis presented in Fig. 4.

With respect to when the adult-type locus duplicated to produce separate δ and β loci, comparisons of coding regions are not informative. Frequent conversions of the δ locus by β in primates, rabbits, rodents, and artiodactyls (2, 5, 29) encompassed coding portions of the two loci. The unconverted noncoding sequences, however, reveal clear homology in δ to δ and β to β interordinal comparisons but low levels of homology in δ to β comparisons. That opossums have only a single adult-type β-globin with no homology between its noncoding regions and those of eutherian δ and β genes suggests that the duplication giving rise to δ- and β-globins occurred in the early eutherians after their divergence from Metatheria.

Evolutionary Implications of Duplicated β-Globin Genes. Our results provide evidence that two progenitors of hemoglobin ε-, γ-, η-, δ-, and β-chain gene lines existed at the time of the metatherian–eutherian split and that these two progenitor genes had already come under different developmental controls. Our phylogenetic reconstruction (Fig. 4) places six nonsynonymous substitutions on the nascent adult hemoglobin β-chain gene line and two nonsynonymous substitutions on the nascent embryonic (ε-like) gene line. Of the hypothesized amino acid changes (in adult β: aspartic acid → glutamic acid, asparagine → aspartic acid, isoleucine → valine, methionine → leucine, isoleucine → valine, and alanine → aspartic acid at positions 43, 47, 54, 78, 111, and 116; in embryonic ε: leucine → phenylalanine and alanine → serine at positions 3 and 52), perhaps the most significant was the replacement of leucine by phenylalanine at position 3 of the embryonic chain. The larger R group of phenylalanine possibly intruded into the space where diphosphoglycerate fits and thus reduced the affinity of the binding site for its ligand. The effect of reduced diphosphoglycerate binding would be to increase the oxygen affinity of embryonic hemoglobin over that of maternal hemoglobin, thereby favoring the transport of oxygen from mother to embryo. Adaptive significance of this change is suggested by the conservation of phenylalanine in opossum ε^M and eutherian ε, γ, and η chains, whereas leucine has been retained by eutherian and opossum adult β-globin chains. With respect to the six hypothesized amino acid changes in the stem of the adult branch, all are found as variants in present-day mammalian β-chains. Nevertheless, three of the six changes occurred at known functional sites: β43 is an $\alpha_1\beta_2$ contact site, β54 is an interior position, and β116 is an $\alpha_1\beta_2$ contact site (6). Thus, it seems reasonable to suggest that the emerging adult hemoglobin had different functional properties from its predecessor.

As just noted, fewer amino acid changes occurred in the progenitor of globins expressed during embryonic stages than in the progenitor of adult β-globins. The numbers of nonsynonymous substitutions on the branches of the most parsimonious tree (Fig. 4) show that this trend continued in metatherians and eutherians. Eutherian ε, the gene consistently expressed in embryonic life, evolved on average at about two-thirds (68%) the rate of eutherian adult β genes, and opossum ε^M evolved at about three-fifths (57%) the rate of opossum β^M. Evidently stabilizing selection acted more pervasively on embryonic ε than on adult β genes. Less room may exist for variation of internal conditions in the embryonic stage of life than in later stages.

In placental mammals, adaptations to extended fetal life within the mother also occurred. At the γ locus, an embryonic hemoglobin was redesigned to be a fetal hemoglobin in simian primates. Convergently, a duplicated β locus became the source of fetal hemoglobin in bovids. Thus, although the expansion of a two-gene β cluster to a five-gene cluster may not have given the early placental mammals any initial advantage over marsupials, the additional loci contributed to the success of later evolutionary experiments.

We thank M. Miyamoto, J. L. Slightom, and R. T. Jones for their comments and suggestions. This study was supported by grants from the National Institutes of Health (R01 HL33940), National Science Foundation (BSR 83-07336), and the Alfred P. Sloan Foundation.

1. 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.
2. Hardison, R. C. (1984) *Mol. Biol. Evol.* **1**, 390–410.

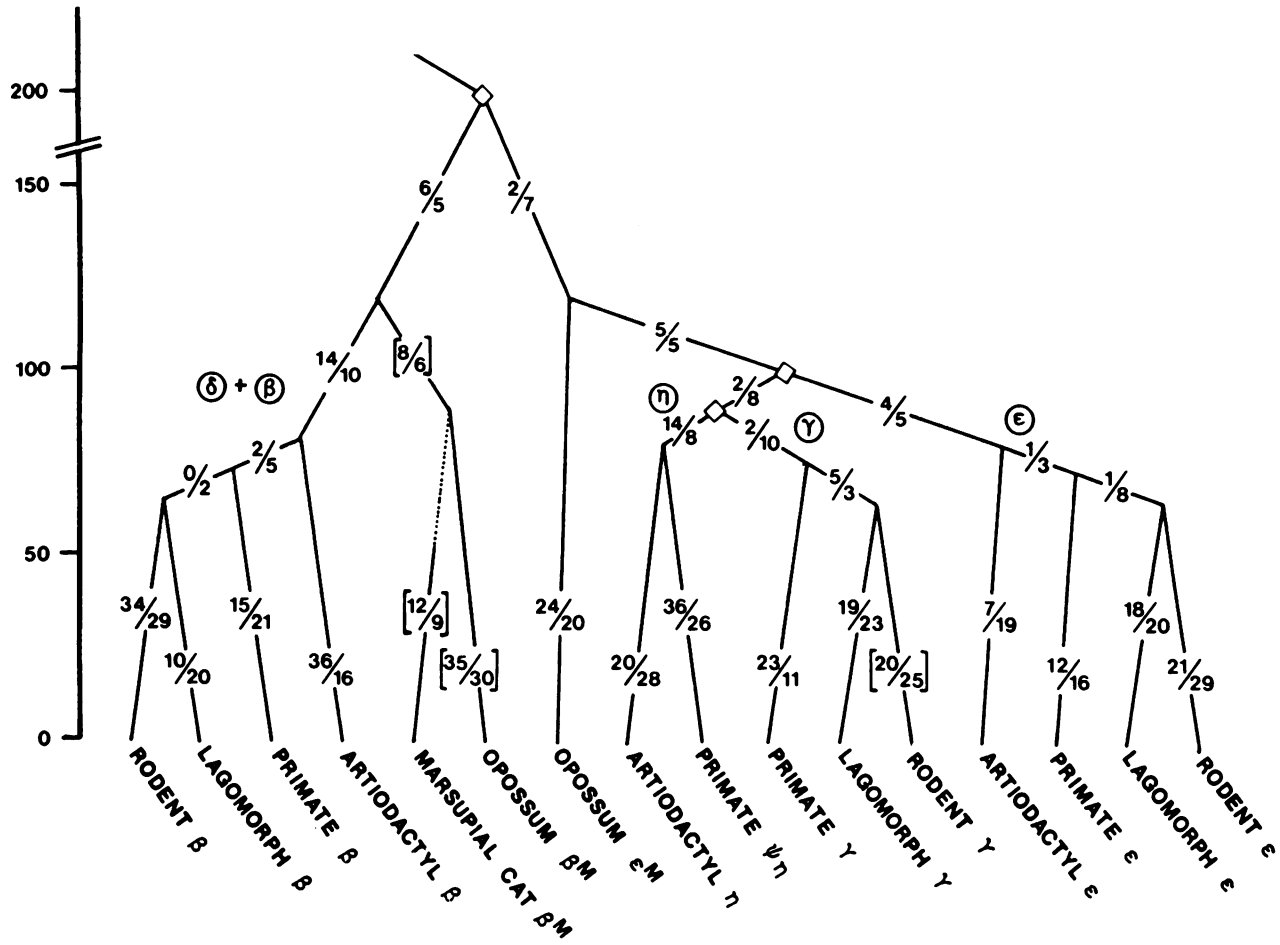

FIG. 4. Sequence relationships as indicated by parsimony analysis. The genealogical arrangement shown here for representative lineages of the β branch of the hemoglobin family is taken from the most parsimonious tree describing the descent of 112 globin genes and pseudogenes, of which 83 were β-globin genes. Nucleotide sequences were aligned to give 444 positions; introns, 5′ and 3′ noncoding DNA, and insertions present only in pseudogenes were not included in the alignment. Sequences that may have been affected by gene conversions or fusions were separated according to proposed boundaries and the separated portions were treated as independent sequences (12, 22–25). Because frequent conversions of δ by β within mammalian orders have eliminated distinctions between the two genes, only the adult β sequences are shown. Link lengths (4, 8) are shown as fractions with the numerator giving the number of amino acid changing substitutions and the denominator giving the number of synonymous substitutions. Terminal branch lengths indicated above are specific for human ε, γ, η, β, rabbit $\beta 4$, $\beta 3$, $\beta 1$, mouse y, βh0, β major, and goat ε^{I}, ε^{II}, β^{A}. Link lengths in parentheses indicate substitutions occurring on branches leading to a partial sequence. Where the placement of a particular mutation was ambiguous, the change was placed on least represented branches. Approximate branching times are obtained by referring to a time bar on the far left.

3. Barrie, P., Jeffreys, A. J. & Scott, A. F. (1981) *J. Mol. Biol.* **149**, 319–336.
4. Goodman, M., Koop, B. F., Czelusniak, J., Weiss, M. L. & Slightom, J. L. (1984) *J. Mol. Biol.* **180**, 803–823.
5. Collins, F. S. & Weissman, S. M. (1984) *Prog. Nucleic Acid Res. Mol. Biol.* **31**, 315–462.
6. Dickerson, R. & Geis, I. (1983) *Hemoglobin: Structure, Function, Evolution and Pathology* (Benjamin/Cummings, Menlo Park, CA).
7. Goodman, M. (1981) *Prog. Biophys. Mol. Biol.* **38**, 105–164.
8. Czelusniak, J., Goodman, M., Hewett-Emmett, D., Weiss, M. L. & Venta, D. J. (1982) *Nature (London)* **298**, 297–300.
9. Stenzel, P., Brimhall, B., Jones, R. T., Black, J. A., McLachlan, A. & Gibson, D. (1979) *J. Biol. Chem.* **254**, 2071–2076.
10. Cifelli, R. L. & Eaton, J. G. (1987) *Nature (London)* **325**, 520–522.
11. Fox, R. C. (1978) *Geol. Ass. Can. Spec. Pap.* **18**, 577–594.
12. Slightom, J. L., Theisen, T. W., Koop, B. F. & Goodman, M. (1987) *J. Biol. Chem.* **262**, 7472–7483.
13. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
14. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 400–560.
15. Nei, M. & Gojobori, T. (1986) *Mol. Biol. Evol.* **3**, 418–426.
16. Ponz, M., Schwartz, E., Ballantine, M. & Surrey, M. (1983) *J. Biol. Chem.* **256**, 11599–11609.
17. Greaves, D. R. & Patient, R. K. (1985) *EMBO J.* **4**, 2617–2626.
18. Shapiro, S. G., Schon, E. A., Townes, T. M. & Lingrel, J. B. (1983) *J. Mol. Biol.* **169**, 31–52.
19. Murphy, W. S., Metcalfe, J., Hoversland, A. S. & Dhindsa, D. S. (1977) *Respir. Physiol.* **29**, 73–80.
20. Wainwright, B. & Hope, R. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 8105–8112.
21. Myers, R. M., Tilly, K. & Maniatis, T. (1986) *Science* **232**, 613–618.
22. Hill, A., Hardies, S. C., Phillips, S. J., Davis, M. G., Hutchinson, C. A., III & Edgell, M. H. (1984) *J. Biol. Chem.* **259**, 1236–1242.
23. Hardison, R. C. (1981) *J. Biol. Chem.* **256**, 11780–11786.
24. Slightom, J. L., Chang, L.-Y. E., Koop, B. F. & Goodman, M. (1985) *Mol. Biol. Evol.* **2**, 370–389.
25. Jeffreys, A. J., Barrie, P., Harris, S., Fawcett, B., Nugent, A. & Boyd, C. (1982) *J. Mol. Biol.* **156**, 487–503.
26. Miyamoto, M. M. & Goodman, M. (1986) *Syst. Zool.* **35**, 230–240.
27. Shoshani, J. (1986) *Mol. Biol. Evol.* **3**, 222–242.
28. Novacek, M. J. (1982) in *Macromolecular Sequences in Systematic and Evolutionary Biology*, ed. Goodman, M. (Plenum, New York) pp. 3–41.
29. Hardies, S. C., Edgell, M. H. & Hutchinson, C. A., III (1984) *J. Biol. Chem.* **256**, 3748–3756.