# New Genes Originated via Multiple Recombinational Pathways in the β-Globin Gene Family of Rodents

*Federico G. Hoffmann,*[1] *Juan C. Opazo,*[2] *and Jay F. Storz*
School of Biological Sciences, University of Nebraska

Species differences in the size or membership composition of multigene families can be attributed to lineage-specific additions of new genes via duplication, losses of genes via deletion or inactivation, and the creation of chimeric genes via domain shuffling or gene fusion. In principle, it should be possible to infer the recombinational pathways responsible for each of these different types of genomic change by conducting detailed comparative analyses of genomic sequence data. Here, we report an attempt to unravel the complex evolutionary history of the β-globin gene family in a taxonomically diverse set of rodent species. The main objectives were: 1) to characterize the genomic structure of the β-globin gene cluster of rodents; 2) to assign orthologous and paralogous relationships among duplicate copies of β-like globin genes; and 3) to infer the specific recombinational pathways responsible for gene duplications, gene deletions, and the creation of chimeric fusion genes. Results of our comparative genomic analyses revealed that variation in gene family size among rodent species is mainly attributable to the differential gain and loss of later expressed β-globin genes via unequal crossing-over. However, two distinct recombinational mechanisms were implicated in the creation of chimeric fusion genes. In muroid rodents, a chimeric γ/ε fusion gene was created by unequal crossing-over between the embryonic ε- and γ-globin genes. Interestingly, this γ/ε fusion gene was generated in the same fashion as the "anti-Lepore" 5′-δ-(β/δ)-β-3′ duplication mutant in humans (the reciprocal exchange product of the pathological hemoglobin Lepore deletion mutant). By contrast, in the house mouse, *Mus musculus*, a chimeric β/δ fusion pseudogene was created by a β-globin → δ-globin gene conversion event. Although the γ/ε and β/δ fusion genes share a similar chimeric gene structure, they originated via completely different recombinational pathways.

## Introduction

The differential gain and loss of genes from homologous gene families represents a potentially important source of functional variation among the genomes of different species (Zhang et al. 2000; Demuth et al. 2006; Hahn, Demuth, and Han 2007; Hahn, Han, and Han 2007; Nozawa and Nei 2007; Hoffmann et al. 2008; Opazo et al. 2008b). Species differences in the size or membership composition of gene families can be attributed to lineage-specific additions of new genes via duplication, losses of genes via deletion or inactivation, and the creation of chimeric genes via domain shuffling or gene fusion (Long et al. 2003; Fan et al. 2008). In principle, each of these different types of genomic change could be produced by similar mechanisms (Shaw and Lupski 2004; Lupski and Stankiewicz 2005; Feuk et al. 2006; Lam and Jeffreys 2006, 2007). For example, unequal crossing-over (nonallelic homologous recombination) between a misaligned pair of linked loci can produce a gene duplication on one daughter chromosome, yielding a recombinant 3-gene haplotype, and a corresponding gene deletion on the other daughter chromosome, yielding a recombinant 1-gene haplotype. Whether this type of cross over event ultimately results in the gain or loss of a gene depends on the evolutionary fates of the two alternative recombinant haplotypes. If the crossover break point occurs within one of the tandemly duplicated genes rather than in the intervening chromosomal region, unequal crossing-over can also produce chimeric fusion genes that contain the 5′ end of one paralog and the 3′ end of the other paralog.

In humans, pathological globin gene deletion mutants such as Hemoglobin (Hb) Lepore, which causes δβ-thalassemia, and Hb Kenya, which causes hereditary persistence of fetal Hb, provide well-characterized examples of how gene duplications, gene deletions, and gene fusions can be generated as reciprocal exchange products of the same recombination event (Dickerson and Geis 1983; Metzenberg et al. 1991; Holloway et al. 2006). For example, crossovers between misaligned copies of the closely linked δ- and β-globin genes result in a solitary δ/β fusion gene on one daughter chromosome and the reciprocal β/δ fusion gene on the other daughter chromosome (Forget 2001; Wood 2001). In the former case, the δ/β fusion gene on the Lepore haplotype is solely responsible for the synthesis of the β-chains of adult Hb. In the latter case, the β/δ fusion gene on the "anti-Lepore" haplotype is flanked by fully functional copies of the parental δ-globin gene on the 5′ side and the parental β-globin gene on the 3′ side. Individuals that are heterozygous for the Hb Lepore deletion haplotype produce adult red blood cells that contain normal $\alpha_2\beta_2$ Hb tetramers (HbA) as well as lesser quantities of $\alpha_2(\delta/\beta)_2$ tetramers (Hb Lepore) that incorporate products of the δ/β fusion gene. The HbA and Hb Lepore isoforms are not present in equal concentrations in red blood cells because the δ/β fusion gene, which is under the control of a δ-type promoter, is transcribed at a much lower rate than the normal β-globin gene. To compensate for the reduced rate of synthesis of adult-type β-chains, the fetal γ-globin genes are upregulated such that heterozygotes for the Hb Lepore deletion haplotype typically only suffer from mild anemia. This compensatory overproduction of fetal Hb is the only reason that the Hb Lepore deletion mutant is not lethal in homozygous condition. In contrast to the complications associated with inheritance of the Hb Lepore deletion haplotype, individuals that inherit the reciprocal product of the same crossover event, the 5′-δ-(δ/β)-β-3′ anti-Lepore haplotype, suffer no pathological effects because they retain a transcriptionally active β-globin gene in addition to the more weakly expressed δ and δ/β genes.

[1] Present address: Instituto Carlos Chagas—ICC—Fiocruz, Curitiba, Brazil.
[2] Present address: Instituto de Ecologia y Evolucion, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile.

Key words: chimeric fusion genes, gene conversion, gene duplication, gene family evolution, globin genes, hemoglobin.

E-mail: jstorz2@unl.edu.

The Hb Lepore deletion haplotype is typically only found at high frequency in geographic regions where malaria was or still is prevalent (Loukopoulos and Kollia 2001).

Thanks to recent advances in the use of single DNA molecule methods that make it possible to detect de novo copy number changes in haploid gametes, we now have a more detailed understanding of the recombinational pathways that have produced the globin gene duplications and deletions that are segregating in the human gene pool (Schneider et al. 2002; Holloway et al. 2006; Lam and Jeffreys 2006, 2007). It is possible that the same recombinational pathways have given rise to the extensive variation in size and membership composition of the globin gene families among different mammalian species (Hardies et al. 1984; Hill et al. 1984; Hardison 1991, 2001; Hoffmann et al. 2008; Opazo et al. 2008b; Storz et al. 2008). Results of sperm-typing experiments indicate that unequal crossing-over represents the primary recombinational pathway for producing copy number changes in the α- and β-globin gene clusters of humans (Holloway et al. 2006; Lam and Jeffreys 2006, 2007). It remains to be seen whether this same mechanism of nonallelic homologous recombination has been responsible for generating species differences in copy number over longer periods of evolutionary time. Gene deletions can also be produced by nonreciprocal exchange processes such as single-strand annealing during DNA break repair (Fishmanlobell et al. 1992). Likewise, gene duplications can be produced by retroposition (Long et al. 2003) or by gene conversion when one preexisting gene is completely overwritten by a paralogous sequence that remains unaltered on the donor chromosome (Tagle et al. 1991). In principle, the evolutionary outcomes of these reciprocal and nonreciprocal recombinational exchanges can be distinguished by conducting detailed comparative analyses of genomic sequence data (Hardison and Miller 1993).

The β-globin gene family of mammals is one of the most intensively studied multigene families from the standpoint of molecular genetics and phylogenetic history (Goodman et al. 1987; Hardison 1991, 2001). The β-globin gene family therefore represents an excellent model system for investigating mechanisms of genome evolution. The β-globin gene cluster of the house mouse, *Mus musculus*, has played an especially prominent role in structural, functional, and comparative genomics (Hardies et al. 1984; Hill et al. 1984; Shehee et al. 1989; Moon and Ley 1990; Hardison and Miller 1993). Early comparative studies of the β-globin gene clusters in human, goat, rabbit, and *Mus* revealed variation in gene copy number among these taxa and inspired two alternative hypotheses to explain the apparently complex evolutionary history of this gene family (Hardies et al. 1984; Hill et al. 1984; Hardison and Miller 1993). The alternative hypotheses invoked different mechanisms to explain evolutionary changes in copy number and the origin of chimeric fusion genes. One especially powerful approach that can be used to reconstruct pathways of gene family evolution is to conduct a comparative analysis of complete genomic sequence contigs from a set of related species that span a broad range of divergence times. Here, we report an attempt to unravel the complex evolutionary history of the β-globin gene family in a taxonomically diverse set of rodent species. The main objectives were 1) to characterize the genomic structure of the β-globin gene cluster of rodents, 2) to assign orthologous and paralogous relationships among duplicate copies of β-like globin genes, and 3) to infer the specific recombinational pathways responsible for gene duplications, gene deletions, and the creation of chimeric fusion genes.

## Material and Methods
### Nomenclature for β-like Globin Genes

Following the nomenclature of Aguileta et al. (2006), we refer to the ε-, γ-, η-, δ-, and β-globin genes as HBE, HBG, HBH, HBD, and HBB, respectively. Pseudogenes are indicated by a "ps" suffix. Because mammalian β-globin genes have undergone multiple rounds of duplication that have resulted in tandemly repeated sets of paralogous gene copies, we index each duplicated gene with the symbol T followed by a number that corresponds to the linkage order in the 5′ to 3′ orientation. For example, the β-like globin genes of *M. musculus* are arranged in the following linkage order on Chromosome 7: 5′-HBE-T1 ($=\varepsilon^y$), HBE-T2 ($=\beta h0$), HBG-T1 ($=\beta h1$), HBD-T1 ($=\beta h2$), HBD-T2 ($=\beta h3$), HBB-T1 ($=\beta 1$), HBB-T2 ($=\beta 2$)-3′ (cf., Hardies et al. 1984; Hill et al. 1984; Hardison and Miller 1993).

### Functional Annotation of Genomic Sequences

We obtained genomic sequences that spanned the entire β-globin gene cluster in seven rodent taxa: Norway rat, *Rattus norvegicus* (family Muridae); house mouse, *M. musculus* (strains C57BL/6 and BALB/c; family Muridae); deer mouse, *Peromyscus maniculatus* (family Cricetidae); white-footed mouse, *Peromyscus leucopus* (family Cricetidae); guinea pig, *Cavia porcellus* (family Caviidae); and thirteen-lined ground squirrel, *Spermophilus tridecemlineatus* (family Sciuridae). For comparative purposes, we also included genomic sequences from human (*Homo sapiens*, order Primates) and rabbit (*Oryctolagus cuniculus*, order Lagomorpha) as outgroups. This sample of genomic sequences included representatives of three suborders of rodents. The suborder Myomorpha was represented by *Mus*, *Rattus*, and *Peromyscus*, the suborder Hystricomorpha was represented by *Cavia*, and the suborder Sciuromorpha was represented by *Spermophilus*. We analyzed publicly available genomic sequence data in the case of *Rattus* (GenBank accession number NC_005100), *Mus* (NC_000067 for the C57BL/6 strain and NT_095534 for the BALB/c strain), *Cavia* (AC186241), *Spermophilus* (AC204819), human (NG_000007), and rabbit (M18818). In the case of the two *Peromyscus* species, we isolated and characterized genomic contigs spanning the entire β-globin gene cluster by screening a bacterial artificial chromosome (BAC) library, as described below.

We annotated the two *Peromyscus* BAC sequences and the other publicly available genomic sequences by using GENSCAN (Burge and Karlin 1997) and by comparing known exon sequences to genomic contigs using the program Blast2 sequences (version 2.2; Tatusova and

Madden 1999). We also used database records to annotate the *Mus* and human genome sequences. For all species, genomic sequences were masked using RepeatMasker (www.repeatmasker.org) and sequence alignments were conducted using PipMaker (Schwartz et al. 2000), Multipipmaker (Schwartz et al. 2003), and Mulan (Ovcharenko et al. 2005).

### BAC Isolation, Sequencing, and Assembly

To isolate and characterize the β-globin gene clusters of *P. maniculatus* and *P. leucopus*, we screened species-specific BAC libraries using one labeled probe designed from the embryonic HBE gene and another labeled probe designed from the adult HBB gene. We designed probes for HBE and HBB because these two genes are located at the 5′ and 3′ ends of the gene cluster in all mammals studied to date. Both probes were designed using an alignment of orthologous coding sequences from *Mus*, *Rattus*, *Cavia*, rabbit, and human. After the hybridization screening, we end sequenced probe-positive BAC clones from each of the two libraries to identify inserts that spanned the entire β-globin gene cluster. A 3- to 4-kb insert high-copy shotgun library was then constructed for each BAC clone selected for full-insert sequencing. The subsequent shotgun sequencing and assembly steps were conducted according to methods described in Storz et al. (2008).

### Phylogeny Estimation

We explored phylogenetic relationships of β-like globin genes at several levels. In all cases, sequences were aligned using ClustalX (Thompson et al. 1997). To reconstruct the history of gene duplication, gene deletion, and gene fusion in the β-globin gene cluster of rodents, we performed separate analyses on coding regions and flanking chromosomal regions. In all cases, we used the human and rabbit β-like globin genes as outgroups. Because conversion tracts are typically restricted to the coding regions of globin genes (Erhart et al. 1985; Hardison and Gelinas 1986; Hardison and Miller 1993; Storz, Baze, et al. 2007; Storz, Sabatino, et al. 2007; Storz et al. 2008; Hoffmann et al. 2008), it is possible to infer orthologous relationships between duplicated genes by examining flanking sequence that lies outside of conversion tracts. Accordingly, phylogeny reconstructions of the β-like globin genes were based on different partitions of a multispecies sequence alignment that started 1 kb upstream of the start codon and ended 1 kb downstream of the stop codon. In all cases, we inferred phylogenetic relationships using Bayesian and maximum likelihood approaches. Bayesian estimation of phylogenies were conducted in MrBayes version 3.1.2 (Ronquist and Huelsenbeck 2003), running four simultaneous chains for $2 \times 10^6$ generations, sampling trees every 1,000 generations, and using default priors. We used a general time-reversible model of nucleotide substitution (Rodriguez et al. 1990) in which rate variation followed a discrete gamma distribution (GTR + Γ). We assessed convergence by measuring the standard deviation of the split frequency among parallel chains. Chains were considered to have converged once the average split frequency was lower than 0.01. We summarized results with a majority-rule consensus of 1,500 trees collected after convergence was reached; trees collected before chains reached convergence were discarded. Maximum likelihood searches were conducted in Treefinder (version April 2008, Jobb et al. 2004). For each data partition, we selected the best fitting model of nucleotide substitution using the Bayesian Information Criterion model selection routine in Treefinder. We evaluated support for the nodes with 1,000 bootstrap pseudoreplicates.

## Results and Discussion
### Genomic Structure of the β-Globin Gene Cluster in Rodents

We isolated, subcloned, and sequenced two BAC clones that contained the entire β-globin gene clusters of *P. maniculatus* and *P. leucopus* (GenBank accession numbers EU204642 and EU559333, respectively). The *P. maniculatus* contig was 181,092 bp in length and the *P. leucopus* contig was 179,761 bp in length. From the start codon of HBE to the stop codon of HBB-T2, the β-globin gene cluster spanned 34.4 kb in *P. maniculatus* and 40.2 kb in *P. leucopus*, and both sequence contigs correspond to nucleotide positions 103686347–103727028 of *Mus* Chromosome 7.

The β-globin gene cluster of mammals contains a set of developmentally regulated genes that are arranged in their temporal order of expression (Collins and Weissman 1984; Hardison 1991, 1998, 2001). The β-globin gene clusters of all rodent taxa in our study conformed to this general arrangement: in each case, the 5′ end of the cluster was delimited by a single embryonic HBE gene and the 3′ end of the cluster was delimited by one or more copies of the later expressed HBB gene (fig. 1).

Consistent with genomic studies of other mammals (Bulger et al. 2000), the β-globin gene cluster was flanked by olfactory receptor genes on both sides. In all rodents, we found an ortholog of the *Mus* Olr66 gene upstream of the HBE gene at the 5′ end of the cluster, and in *Rattus* and *Peromyscus*, we found an ortholog of the *Mus* Olr67 gene downstream of the last HBB gene at the 3′ end of the cluster. In all rodent species studied to date, the HBE and HBG genes are expressed in embryonic erythroid cells derived from the yolk sac and the HBB genes are expressed in fetal and adult erythroid cells (Hardison 2001). The HBE and HBG genes of contemporary rodents originated via duplication of a proto ε-globin gene, and the HBDps and HBB genes originated via duplication of a proto β-globin gene. In both cases, the duplication events occurred in the ancestor of eutherian mammals after the divergence from marsupials (Goodman et al. 1984; Koop and Goodman 1988; Cooper et al. 1996; Opazo et al. 2008a). Following the convention of Hill et al. (1984) and Hardies et al. (1984), we refer to the 3′ end of the gene cluster (containing the embryonically expressed HBE and HBG genes) as the "nonadult" portion and the 5′ end of the cluster (containing the HBDps pseudogenes and later expressed HBB genes) as the "adult" portion.
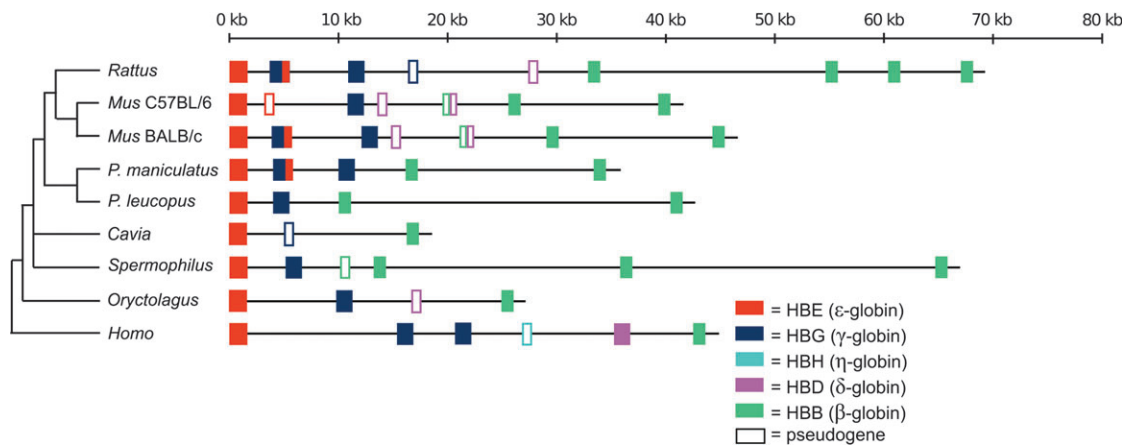
FIG. 1.—Genomic structure of the β-globin cluster in rodents and two outgroup species—human and rabbit. The depicted phylogenetic relationships are based on a loose consensus of recent studies (Murphy et al. 2001; Steppan et al. 2004; Prasad et al. 2008). The orientation of the clusters is from 5′ (on the left) to 3′ (on the right).

Comparison of the β-globin gene clusters of the seven rodent taxa in our study revealed considerable variation in gene copy number, especially among the adult HBB genes at the 3′ end of the cluster. Whereas the gene cluster of *Cavia* contained a single embryonic HBE gene and a single adult HBB gene, the gene cluster of *Rattus* contained seven putatively functional genes, four of which were HBB paralogs (fig. 1).

Pairwise sequence comparisons between the human β-globin gene cluster and each of the different rodent taxa revealed high levels of sequence conservation in the chromosomal region upstream of the HBE gene (supplementary fig. S1, Supplementary Material online). This chromosomal region contains the locus control region (LCR), a complex array of *cis*-acting control elements spanning ~20 kb of sequence that regulates transcription of all β-like globin genes in the cluster (Grosveld et al. 1987; Hardison et al. 1997). In each pairwise human–rodent comparison, the chromosomal region extending from the LCR to the HBE gene represents the only region of extended co-linear sequence matches. Downstream of the HBE gene, sequence matches become shorter and are mostly restricted to exons of the β-like globin genes.

## Ancestral State of the Rodent β-Globin Gene Cluster

The human β-globin gene cluster contains six genes: 5′-HBE, HBG-T1, HBG-T2, HBHps, HBD, and HBB-3′ (Fritsch et al. 1980), whereas the rabbit (the outgroup species that is most closely related to rodents) contains four genes: 5′-HBE, HBG, HBDps, and HBB-3′ (Margot et al. 1989; fig 1). Comparisons with the human β-globin gene cluster show that there is no ortholog of the human HBHps pseudogene in rabbit or in any of the rodents included in our analysis, consistent with the findings of Goodman et al. (1984). The most parsimonious explanation is that the HBH gene was deleted in the common ancestor of rodents and lagomorphs. According to the principle of parsimony, the ancestral rodent gene cluster most likely

contained four β-like globin genes in the following linkage order: 5′-HBE, HBG, HBDps, HBB-3′, similar to the gene cluster of the rabbit (Collins and Weissman 1984; Hardison 1984, 1991). Accordingly, we used the rabbit gene cluster as our reference for the purpose of reconstructing gene gains, losses, and fusions during the evolutionary history of the rodent β-globin gene cluster.

## Reconstructing Pathways of β-Globin Gene Family Evolution

Relative to the inferred 4-gene structure of the ancestral β-globin gene cluster, the sample of rodent species included in our study exhibited an evolutionary trend of increasing gene number. With the exception of *Cavia*, all rodent taxa in our study have added 1–3 HBB paralogs at the 3′ end of the cluster, and muroid rodents have added a γ/ε chimeric fusion gene that is located between the parental HBE and HBG genes at the 5′ end of the cluster (fig. 1). If this chimeric fusion gene does in fact have a single origin in the common ancestor of muroid rodents, then it appears to have been secondarily lost in *P. leucopus* and in the C57BL/6 strain of *M. musculus*. Because this locus should be named HBE-T2 on the basis of positional homology, we henceforth use the name "HBE(γ/ε)-T2" to indicate that the 5′ portion is γ like and the 3′ portion is ε like. In *Mus*, an additional chimeric β/δ pseudogene is located between the parental HBD and HBB genes (fig. 1). Because this locus should be named "HBD-T2ps" on the basis of positional homology, we henceforth use the name "HBD(β/δ)-T2ps" to indicate that the 5′ portion is β like and the 3′ portion is δ like.

To assess whether any of the above-mentioned gene gains and gene fusions occurred independently as lineage-specific events, we used phylogenetic reconstructions in combination with pairwise analyses of sequence similarity. Phylogeny reconstructions of the nonadult genes (HBE, HBE(γ/ε)-T2, and HBG) and the later expressed HBB genes are shown in figures 2 and
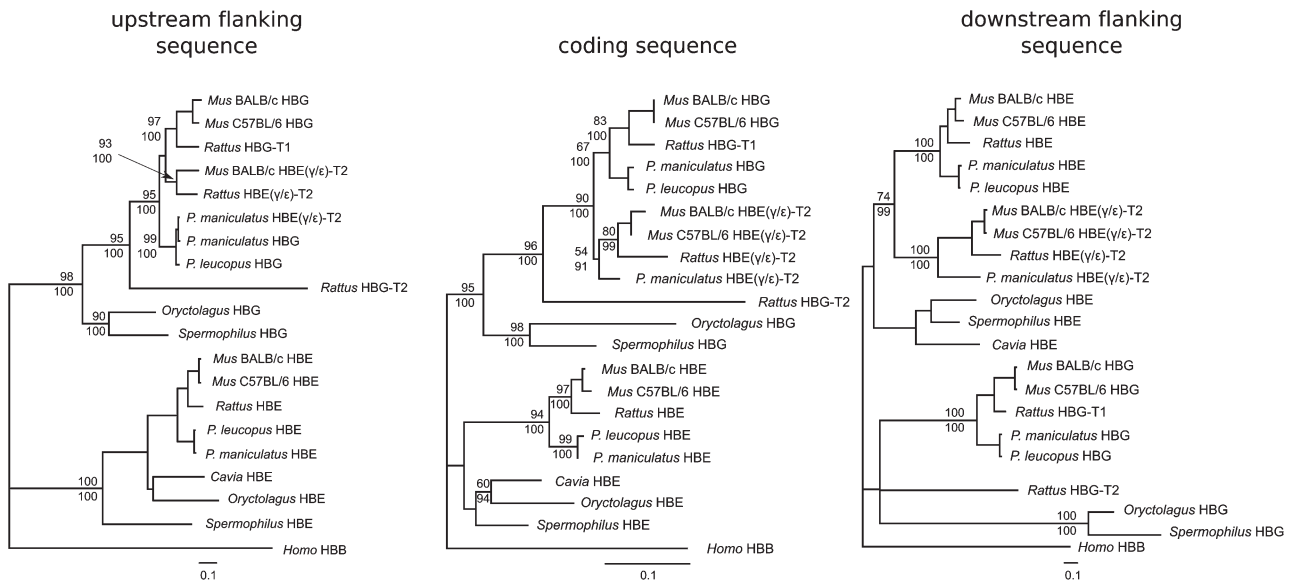
FIG. 2.—Maximum likelihood phylograms depicting relationships among embryonic β-like globin genes of rodents. The phylogeny reconstructions were based on 1 kb of 5′-flanking sequence (left column), coding sequence (center column), and 1 kb of 3′-flanking sequence (right column). Note that the guinea pig HBGps pseudogene is excluded from the analyses because the 3′ portion of the coding region has been deleted. Measures of support for the relevant nodes are presented as bootstrap values (above the nodes) and as Bayesian posterior probabilities (below the nodes).

3, respectively. Separate reconstructions of the HBE and HBG paralogs (with rabbit as outgroup) are shown in supplementary fig. S2 (Supplementary Material online). The high divergence among the HBD-like genes did not permit robust alignments of flanking sequences, so orthologous relationships among the HBD-like genes were resolved using pairwise dot plots and Blast comparisons. We used the phylogeny reconstructions and patterns of pairwise sequence similarity to infer the set of orthologous relationships for all β-like globin genes, as illustrated in figure 4.

## Nonadult Portion of the β-Globin Gene Cluster (HBE and HBG Genes)

Phylogeny reconstructions based on upstream sequence and coding sequence grouped HBE(γ/ε)-T2 fusion genes with the HBG genes, whereas reconstructions based on downstream sequence grouped the fusion genes with the HBE genes (fig. 2). The fact that the upstream flanking sequence of the HBE(γ/ε)-T2 fusion gene is HBE like in character, whereas the downstream flanking sequence is HBG like in character indicates that the fusion gene originated via
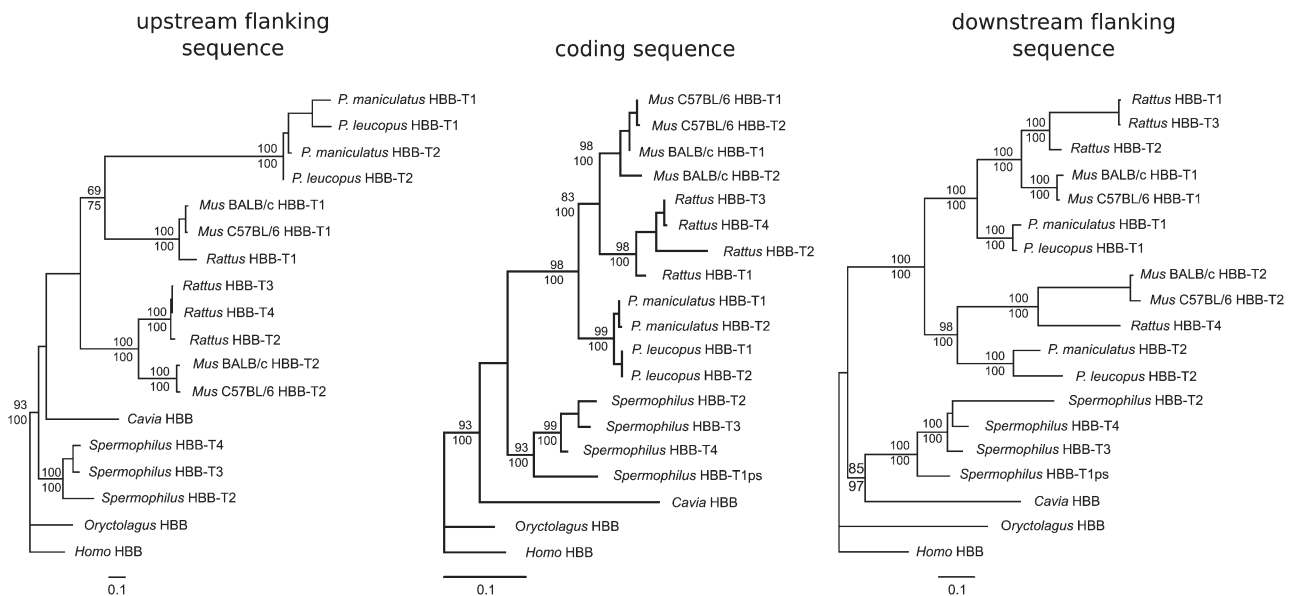


FIG. 3.—Maximum likelihood phylograms depicting relationships among HBB genes of rodents. The phylogeny reconstructions were based on 1 kb of 5′-flanking sequence (left column), coding sequence (center column), and 1 kb of 3′-flanking sequence (right column). Measures of support for the relevant nodes are presented as bootstrap values (above the nodes) and as Bayesian posterior probabilities (below the nodes).
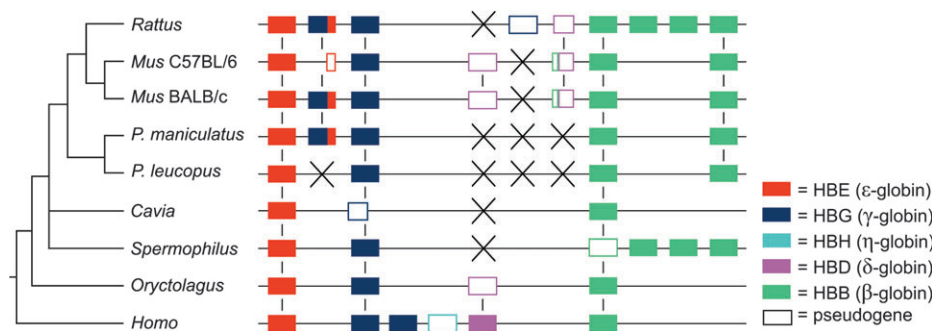
FIG. 4.—Orthologous relationships among β-like globin genes of rodents, rabbit, and human, as inferred from comparative analyses of 5′- and 3′-flanking sequences. Solid boxes denote putatively functional genes, open boxes denote pseudogenes, and crosses denote gene deletions. Vertical lines are used to indicate 1:1 orthologous relationships.

unequal crossing-over and that the recombination break point interrupted the coding region of the misaligned HBE and HBG genes. Phylogeny reconstructions of coding sequences in addition to upstream and downstream flanking sequences grouped HBE and HBG genes into two reciprocally monophyletic groups. For the most part, the true set of species relationships was recovered independently in each of the two paralogous clades. The few discrepancies (e.g., the phylogenetic placement of rabbit relative to *Spermophilus* and *Cavia* in the trees based on upstream flanking sequence and coding sequence) can be explained by ambiguities in the alignment of noncoding sequence. The phylogeny reconstructions clearly indicate that the single-copy HBE genes of rodents, rabbit, and human are 1:1 orthologs. Likewise, phylogeny reconstructions indicate that the 5′ HBG genes of these same taxa are also 1:1 orthologs (figs. 2 and 4). Ambiguities in the alignment of flanking sequence from different paralogs do not appear to have had a major effect on the phylogeny reconstructions as the same sets of relationships among rodent sequences were recovered in phylogeny reconstructions that treated the HBE and HBG paralogs separately (supplementary fig. S2, Supplementary Material online).

Because *Rattus* possesses an additional copy of HBG (HBG-T2ps) relative to the rodent ancestor, it would seem reasonable to conclude that this pseudogene represents the product of a duplication event that was specific to the *Rattus* lineage. If this were the case, the HBG-T1 and HBG-T2ps sequences of *Rattus* should be more similar to one another than they are to the single-copy HBG pro-orthologs of the other rodent species. Contrary to this expectation, phylogeny reconstructions based on coding sequence as well as flanking sequence show that the *Rattus* HBG-T2ps pseudogene is basal to a monophyletic clade that contains the single-copy HBG genes of other muroid rodents as well as the *Rattus* HBG-T1 paralog (fig. 2). This indicates that the *Rattus* HBG-T2ps pseudogene is the product of a relatively ancient duplication event that predates the divergence of murid and cricetid rodents. Thus, orthologs of the *Rattus* HBG-T2ps pseudogene must have been secondarily deleted in *Mus* and *Peromyscus* (fig. 4).

## Adult Portion of the β-Globin Gene Cluster (HBD and HBB Genes)

The HBD-T1ps pseudogene of *Mus* and the HBDps pseudogene of *Rattus* both exhibit clear affinities to HBDps of rabbit (fig. 5). Surprisingly, however, a pairwise comparison of sequence similarity between *Rattus* and *Mus* (fig. 6) revealed that the HBDps pseudogene of *Rattus* is not a 1:1
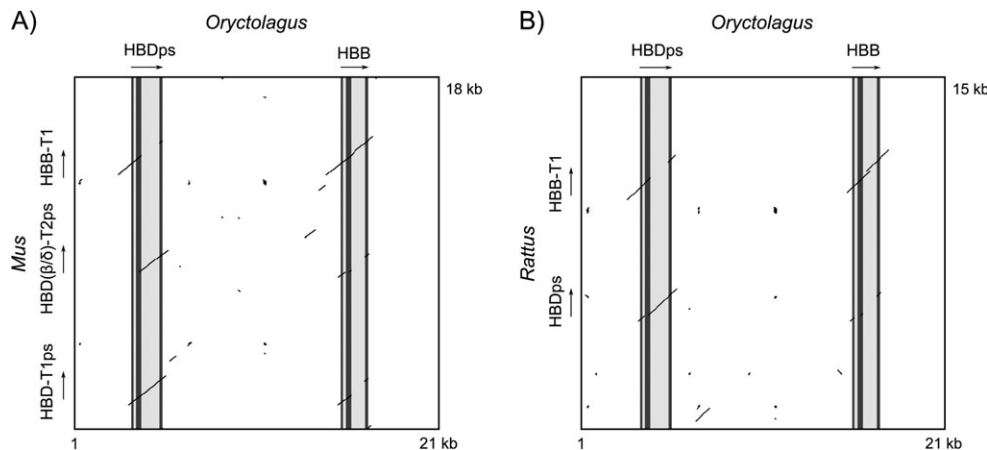


FIG. 5.—Dot plots of sequence similarity between the HBD and HBB genes of rabbit, *Oryctolagus* (horizontal axis) and their presumptive orthologs in *Mus* (A, vertical axis) and *Rattus* (B, vertical axis). The pairwise comparisons were based on a chromosomal region that contained the HBB-T1 genes and HBD-like pseudogenes of both rodent species.
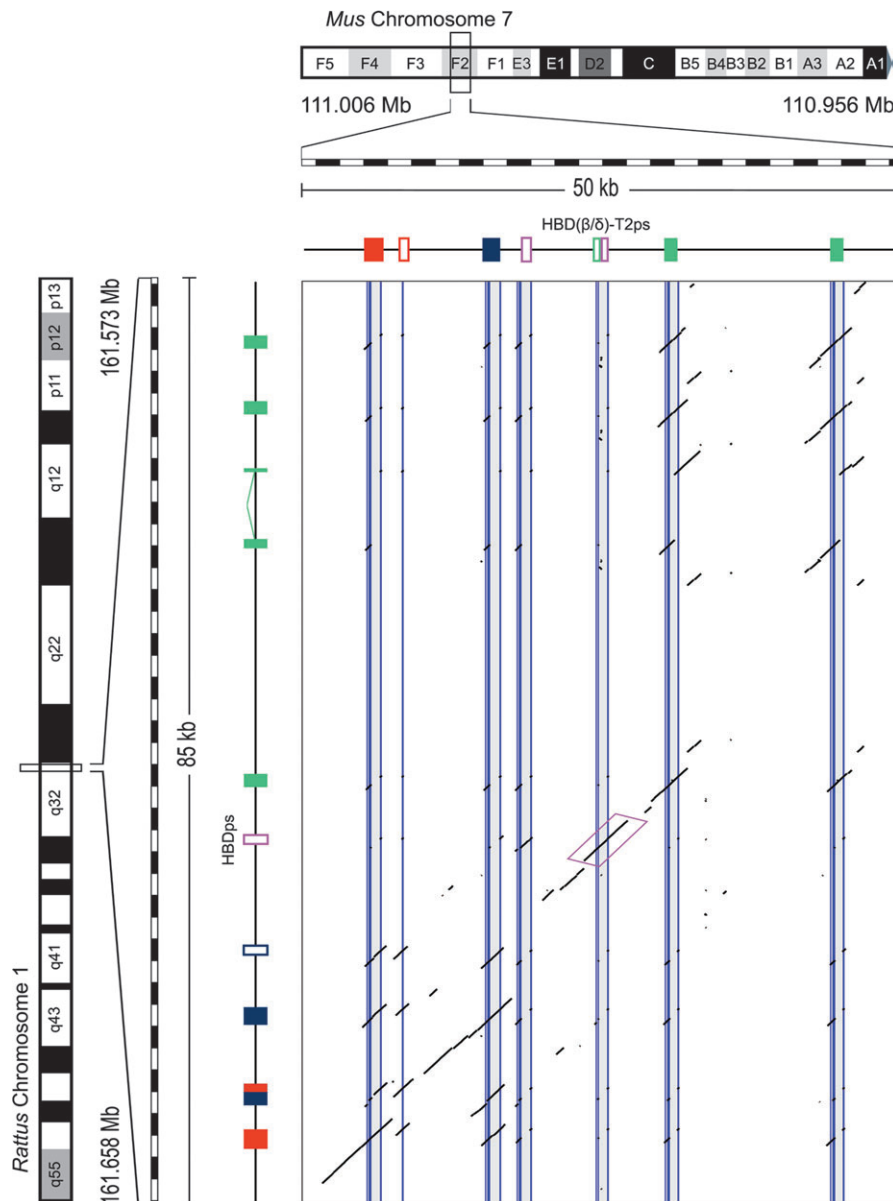
FIG. 6.—Dot plot of a chromosomal fragment spanning the β-globin gene cluster on Chromosome 7 of house mouse (*Mus musculus*, strain C57BL/6) and the syntenic region on Chromosome 1 of rat (*Rattus norvegicus*), with masked repeats. The polygon with magenta outline encloses a locally alignable chromosomal region that provides evidence for a 1:1 orthologous relationship between HBDps of *Rattus* and HBD(β/δ)-T2ps of *Mus*. Specifically, the pattern of pairwise sequence matches suggests that HBDps of *Rattus* is a 1:1 ortholog of the ancestral HBD-T2 gene of *Mus*, the 5′ end of which was later overwritten by a short conversion tract derived from HBB-T1 sometime after the divergence between *Rattus* and *Mus*.

ortholog of the HBD-T1ps pseudogene that has been retained in the gene cluster of *Mus*. The pattern of pairwise sequence matches across the coding region and flanking regions shows that the HBDps pseudogene of *Rattus* is orthologous to the gene that became the HBD(β/δ)-T2ps fusion gene in *Mus* (fig. 6). We infer that this chimeric fusion gene was created by conversion of an ancestral HBD-T2ps pseudogene by the upstream HBB-T1 gene in the *Mus* lineage. Thus, HBDps of *Rattus* reflects the ancestral state of the HBD(β/δ)-T2ps fusion gene of *Mus* before the 5′ end was overwritten by the HBB-T1 → HBD-T2ps conversion event. We conclude that the fusion gene was created by gene conversion rather than unequal crossing-over because

the upstream flanking sequence of *Mus* HBD(β/δ)-T2ps provides a much better match to the upstream flanking sequence of *Rattus* HBDps than to the upstream flanking sequence of the adjacent HBB paralog (fig. 6). If the HBD(β/δ)-T2ps fusion gene were an anti-Lepore product of unequal crossing-over, the 5′-flanking sequence would be HBB like in character and would therefore match the orthologous 5′-flanking sequence of the *Rattus* HBB-T1 gene.

In phylogeny reconstructions of HBB genes based on coding sequence, paralogs from the same species often clustered together to the exclusion of their presumptive orthologs in other species (fig. 3). The within-species monophyly of tandemly duplicated genes is a hallmark of

concerted evolution and, in the case of the 5′ and 3′ HBB paralogs of muroid rodents, indicates that interparalog gene conversion has obscured the true evolutionary history of gene duplication and species divergence. In contrast to the phylogenies based on coding sequence, phylogenies based on flanking sequence of the 5′ and 3′ HBB genes generally recovered the true species relationships within each paralogous clade. This pattern confirms that gene conversion tracts are primarily restricted to coding regions, consistent with findings from previous studies of rodent globin genes (Erhart et al. 1985; Storz, Baze, et al. 2007; Storz, Sabatino, et al. 2007; Storz et al. 2008). The two HBB paralogs of *Peromyscus* represent the only exception to this general pattern. In the case of the two *Peromyscus* species, phylogenies based on coding sequence show that the HBB-T1 and HBB-T2 paralogs from the same species are more similar to each other than they are to their presumptive orthologs in the other congeneric species, whereas phylogenies based on downstream flanking sequence show that the HBB-T1 genes of *P. maniculatus* and *P. leucopus* are sister to one another and likewise for the HBB-T2 genes (fig. 3). In the phylogenies based on upstream flanking sequence, however, the HBB-T2 sequences of both species are nested within the clade of HBB-T1 sequences from other muroid rodents. This indicates that in both *Peromyscus* species a conversion tract derived from HBB-T1 extends partway into the 5′-flanking region of HBB-T2.

The HBB-T1 and HBB-T2 paralogs from the two different strains of *M. musculus* show the same topological incongruence between the phylogeny based on coding sequence and the phylogenies based on upstream and downstream flanking sequence (fig. 3). However, the HBB-T1- and HBB-T2-coding sequences are not reciprocally monophyletic within each strain. This is due to the fact that the BALB/c and C57BL/6 strains carry two alternative haplotypes that have different histories of gene conversion between the closely linked HBB-T1 and HBB-T2 paralogs (Storz, Baze, et al. 2007). The C57BL/6 ($Hbb^s$) haplotype carries two HBB-T1 and HBB-T2 paralogs that are identical in sequence, whereas the BALB/c ($Hbb^d$) haplotype carries two highly divergent HBB-T1 and HBB-T2 paralogs that are distinguished by nine amino acid substitutions. As a result of interparalog gene conversion on the $Hbb^s$ haplotype, the HBB-T1 sequence of BALB/c is more similar to both paralogs of C57BL/6 than to the HBB-T2 paralog of BALB/c (Fig. 3).

Phylogeny reconstructions based on flanking sequence clearly show that the 5′ HBB genes of the rodents, rabbit, and human are 1:1 orthologs. Likewise, phylogeny reconstructions based on flanking sequence clearly show that the 3′ HBB genes of all muroid rodents are 1:1 orthologs (fig. 3). By contrast, the HBB-T2, HBB-T3, and HBB-T4 paralogs of *Spermophilus* cluster together in phylogenies based on coding sequence and in those based on upstream and downstream flanking sequence (fig. 3). This suggests that these three genes have originated via lineage-specific duplication events. In *Spermophilus*, the duplication that gave rise to the HBB-T4 gene involved a ~10-kb block that contained the ancestral HBB-T3 gene. By contrast, the duplication that gave rise to the HBB-T2 gene

involved a much smaller block that did not extend much beyond the coding region of the ancestral HBB-T1 gene (supplementary fig. S3, Supplementary Material online). The same phylogenetic pattern indicates that the HBB-T2 and HBB-T3 paralogs of *Rattus* are also attributable to lineage-specific duplications via unequal crossing-over (fig. 3). High rates of recurrent unequal crossing-over also appear to be responsible for HBB copy number polymorphism among different lab strains of *R. norvegicus*, as haplotypes with 3–5 HBB genes have been documented (Stevanovic et al. 1989; Paunesku et al. 1990).

## Pathways Involved in the Addition of New Genes to the β-Globin Gene Family

Cases involving an increase in copy number of tandem gene duplicates are typically attributable to unequal crossing-over (Hoffmann et al. 2008; Storz et al. 2008). This same recombinational mechanism can also produce chimeric fusion genes when the crossover break point interrupts the coding regions of two misaligned tandem duplicates. Alternatively, chimeric genes can also be produced by interparalog gene conversion (Jeffreys et al. 1982; Hardison and Margot 1984; Koop et al. 1989; Tagle et al. 1991; Prychitko et al. 2005). The former mechanism may be more likely to alter expression patterns because it simultaneously produces a change in copy number and a change in the spacing of genes relative to distal *cis*-regulatory elements. Cases in which gene number increases in multiples of two (from two to four, three to six, etc.) are often attributable to the tandem duplication of relatively large chromosomal blocks that contain two or more closely linked genes. For example, the β-globin gene cluster of goats consists of a triplication of a set of four genes: 5′-HBE-T1, HBE-T2, HBB-T1ps, and HBB-T2-3′ (Townes et al. 1984). Similar *en bloc* duplications of β-like globin genes have also been documented in sheep (Garner and Lingrel 1988, 1989) and cows (Schimenti and Duncan 1985).

Previous studies have proposed different hypotheses to explain the complex evolutionary history of gene gains, losses, and fusions in the β-globin gene cluster of rodents (Hardies et al. 1984; Hill et al. 1984; Hardison and Miller 1993). Specifically, the model that was outlined in a pair of companion papers by Hardies et al. (1984) and Hill et al. (1984)—the "HH84" model—and the model outlined by Hardison and Miller (1993)—the "HM93" model—proposed different recombinational pathways to account for changes in copy number and the creation of chimeric fusion genes in the β-globin gene cluster of *M. musculus*. The HH84 and HM93 models are graphically depicted in figure 7*A* and *B*, respectively. Our own model, based on inferences drawn from the comparative genomic analyses described above, is graphically depicted in figure 7*C* for comparison.

With regard to the nonadult portion of the gene cluster, the HH84 model postulates a tandem duplication of HBG (fig. 7*A*, step 1), whereas the HM93 model postulates a similar tandem duplication of HBE (fig. 7*B*, step 1). Although both models invoke interparalog gene conversion to explain the creation of the γ/ε fusion gene in *M. musculus*, they
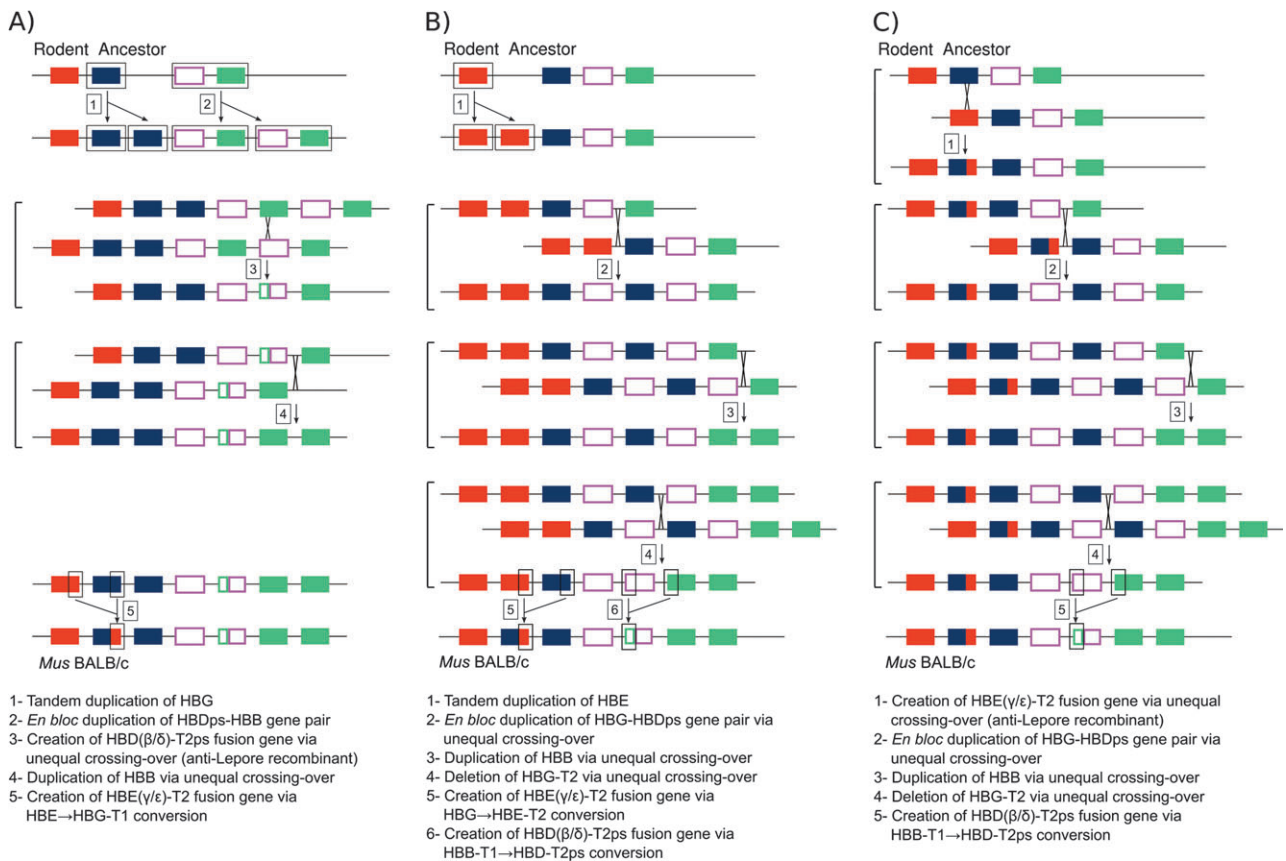
Fig. 7.—Alternative models for the evolution of the mouse β-globin gene cluster. Three alternative pathways are illustrated for the derivation of the *Mus musculus* (BALB/c) β-globin gene cluster from the ancestral gene arrangement in the stem lineage of rodents (see text for details). (*A*) The HH84 model outlined by Hardies et al. (1984) and Hill et al. (1984); (*B*) The HM93 model outlined by Hardison and Miller (1993); and (*C*) Our proposed model based on inferences drawn from a comparative genomic analysis of the β-globin gene cluster in multiple rodent taxa. Note that the HM93 model did not specify whether the creation of the HBD(β/δ)-T2ps fusion gene was attributable to interparalog gene conversion or unequal crossing-over, although the fusion event is depicted as a result of HBB → HBDps gene conversion (panel *B*, step 6).

make different predictions about the directionality of the conversion event and the ancestral identity of the converted gene. The HH84 model postulates that the fusion gene was created by conversion of HBG-T1 by the 3′ end of HBE (fig. 7*A*, step 5), whereas the HM93 model postulates that the fusion gene was created by conversion of HBE-T2 by the 3′ end of HBG (fig. 7*B*, step 5). Results of our analyses do not support either of these proposed scenarios. Instead, as explained above, our phylogeny reconstructions of flanking sequence (fig. 2) indicate that the γ/ε fusion gene was created by unequal crossing-over between the single copy HBE and HBG paralogs that were present in the rodent common ancestor (fig. 7*C*, step 1). This inferred recombination event is similar to the intragenic crossover that is responsible for producing the 5′-δ-(β/δ)-β-3′ anti-Lepore duplication mutant in humans (Metzenberg et al. 1991). In addition to the fact that it can be reconciled with the phylogenetic results, this scenario is also more parsimonious than those invoked by the other models as it requires a single step to account for the discordant phylogenetic affinities of the upstream and downstream flanking regions of the HBE(γ/ε)-T2 fusion gene.

With regard to the adult portion of the gene cluster, the HH84 model postulates an *en bloc* duplication of an HBDps–HBB gene pair (fig. 7*A*, step 2), whereas the HM93 model postulates two successive rounds of unequal crossing-over to explain the addition of HBD-T2ps (along with HBG-T2; fig. 7*B*, step 2) and the addition of HBB-T2 (fig. 7*B*, step 3). According to the HM93 model, the *en bloc* duplication of the HBG-HBDps gene pair is followed by deletion of HBG-T2 in a subsequent round of unequal crossing-over (fig. 7*B*, step 4). This step in the HM93 model explains the observed similarity between the upstream flanking sequence of HBG and the intergenic sequence between HBD-T1ps and HBD(β/δ)-T2ps (Hardison and Miller 1993). Our results are consistent with the predictions of the HM93 model and indicate that each of the postulated copy number changes in the adult portion of the gene cluster must have occurred in the common ancestor of muroid rodents. If an *en bloc* duplication of the HBG-HBDps gene pair occurred in the common ancestor of muroid rodents, then the only surviving descendent of the ancestral HBD-T1 paralog would be the HBD-T1ps pseudogene in *Mus* and the only surviving descendents of the ancestral HBD-T2 paralog would be the HBDps pseudogene of *Rattus* and the HBD(β/δ)-T2ps fusion gene of *Mus* (although only the unconverted HBD-like portion of the fusion gene would accurately reflect the true orthologous relationship).

Likewise, the only surviving descendent of the ancestral HBG-T2 paralog would be HBG-T2ps of *Rattus*. Each of these postulated steps can be reconciled with our inferred set of orthologous relationships among the β-like globin genes (fig. 4).

The HH84 model also invokes a specific recombinational pathway to explain the creation of the HBD(β/δ)-T2ps fusion gene. This model postulates that the HBD(β/δ)-T2ps fusion gene was created by unequal crossing-over between misaligned copies of the recently derived HBB-T1 and HBD-T2 paralogs (fig. 7*A*, step 3). The HM93 model did not specify whether the creation of the HBD(β/δ)-T2ps fusion gene was attributable to interparalog gene conversion or unequal crossing-over (though we took the liberty of depicting the fusion event as the result of an HBB → HBDps gene conversion; Fig. 7*B*, step 6). At face value, the HH84 model appears to be quite plausible because the intragenic cross over between misaligned copies of HBB-T1 and HBD-T2 would produce the observed 5′-HBD-HBD(β/δ)-T2ps-HBB-3′ haplotype in exactly the same manner as the 5′-δ-(β/δ)-β-3′ anti-Lepore haplotype of humans. It seems reasonable to expect that the two superficially similar anti-Lepore gene arrangements would have been produced by the same recombinational mechanism. However, results of our analysis indicate that this is not the case. The anti-Lepore recombinational pathway proposed by the HH84 model makes two key predictions: 1) The upstream flanking sequence of the HBD(β/δ)-T2ps fusion gene should match that of HBB-T1 and 2) the HBDps pseudogene of *Rattus* should be orthologous to the HBD-T1ps pseudogene of *Mus*. Results of our analyses are not consistent with either of these two predictions. First of all, pairwise analyses of sequence similarity between *Mus* and *Rattus* (fig. 6) clearly demonstrate that the 5′-flanking sequence of the HBD(β/δ)-T2ps fusion gene is HBD like in character (not HBB like, as predicted by the HH84 model). Secondly, the same comparison between *Mus* and *Rattus* (fig. 6) clearly demonstrates that *Rattus* HBDps is a 1:1 ortholog of *Mus* HBD(β/δ)-T2ps (not of *Mus* HBD-T1ps, as predicted by the HH84 model). At least where the adult portion of the gene cluster is concerned, our analysis is consistent with the evolutionary pathway proposed by the HH93 model (cf., fig. 7*B* and *C*).

## Conclusion

Results of our comparative genomic analyses demonstrate that several different pathways were involved in the addition of new genes to the β-globin gene family of rodents. During the course of rodent evolution, the addition of new β-like globin genes was mainly attributable to unequal crossing-over. Similarly, the HBE(γ/ε)-T2 fusion gene of muroid rodents was the product of unequal crossing-over between the embryonic HBE and HBG genes. Results of our comparative analysis of sequence variation indicate that this HBE(γ/ε)-T2 fusion gene was generated in the same fashion as the anti-Lepore HBD(β/δ)-T2 mutant in humans (Metzenberg et al. 1991). By contrast, the HBD(β/δ)-T2ps pseudogene of *Mus* derives from a HBB → HBD gene conversion event. Although the HBE(γ/ε)-T2 and

HBD(β/δ)-T2ps loci of *Mus* share a similar chimeric gene structure, the two fusion genes originated via completely different recombinational pathways. Hardies et al. (1984) suggested that HBE(γ/ε)-T2 encodes the β-chain subunits of an "early" embryonic Hb that performs a specialized physiological function during the earliest stages of embryogenesis. If there is some physiological division of labor between the (γ/ε)-chain isoforms and the other prenatally expressed β-chain Hb isoforms, it would be interesting to know whether the specialized functions of the (γ/ε)-chain Hb isoforms are attributable to the unique, chimeric nature of the β-chain polypeptide that has an γ-like N-terminal portion and an ε-like C-terminal portion.

## Supplementary Material

Supplementary figures S1–S3 and all alignments associated with this manuscript are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/). Sequence data from this article have been submitted to GenBank under accession nos. EU204642 and EU559333.

## Literature Cited

Aguileta G, Bielawski JP, Yang ZH. 2006. Proposed standard nomenclature for the alpha- and beta-globin gene families. Genes Genet Syst. 81:367–371.

Bulger M, Bender MA, Hikke van Doorninck J, Wertman B, Farrell CM, Felsenfeld G, Groudine M, Hardison R. 2000. Comparative structural and functional analysis of the olfactory receptor genes flanking the human and mouse β-globin gene clusters. Proc Natl Acad Sci USA. 97: 14560–14565.

Burge C, Karlin S. 1997. Prediction of complete gene structure in human genomic DNA. J Mol Biol. 268:78–94.

Collins FS, Weissman SM. 1984. The molecular genetics of human hemoglobin. Prog Nucleic Acid Res Mol Biol. 31:315–462.

Cooper SJB, Murphy R, Dolman G, Hussey D, Hope RM. 1996. A molecular and evolutionary study of the beta-globin gene family of the Australian marsupial *Sminthopsis crassicaudata*. Mol Biol Evol. 13:1012–1022.

Demuth JP, De Bie T, Stajich JE, Cristianini N, Hahn MW. 2006. The evolution of mammalian gene families. PLoS ONE. 1:e85.

Dickerson RE, Geis I. 1983. Hemoglobin: structure, function, evolution, and pathology. Menlo Park (CA): Benjamin/Cummings Publishing.

Erhart MA, Simons KS, Weaver S. 1985. Evolution of mouse β-globin genes: a recent gene conversion in the Hbb$^s$ haplotype. Mol Biol Evol. 2:304–320.

Fan C, Emerson JJ, Long M. 2008. The origin of new genes. In: Pagel M, Pomiankowski A, editors. Evolutionary genomics and proteomics. Sunderland (MA): Sinauer Associates. p. 27–43.

Feuk L, Carson AR, Scherer SW. 2006. Structural variation in the human genome. Nat Rev Genet. 7:85–97.

Fishmanlobell J, Rudin N, Haber JE. 1992. Two alternative pathways of double-strand break repair that are kinetically separable and independently modulated. Mol Cell Biol. 12: 1292–1303.

Forget BG. 2001. Molecular mechanisms of β thalassemia. In: Steinberg MH, Forget BG, Higgs DR, Nagel RL, editors. Disorders of hemoglobin: genetics, pathophysiology, and clinical management. Cambridge: Cambridge University Press. p. 252–276.

Fritsch E, Lawn R, Maniatis T. 1980. Molecular cloning and characterization of the human beta-like globin gene cluster. Cell. 19:959–972.

Garner KJ, Lingrel JB. 1988. Structural organization of the β-globin locus of B-haplotype sheep. Mol Biol Evol. 5:134–140.

Garner KJ, Lingrel JB. 1989. A comparison of the beta A- and beta B-globin clusters of sheep. J Mol Evol. 28:175–184.

Goodman M, Czelusniak J, Koop BF, Tagle DA, Slightom JL. 1987. Globins: a case study in molecular phylogeny. Cold Spring Harb Symp Quant Biol. 52:875–890.

Goodman M, Koop BF, Czelusniak J, Weiss ML. 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, van Assendelft GB, Greaves D, Kollias G. 1987. Position-independent, high-level expression of the human β-globin gene in transgenic mice. Cell. 51:975–985.

Hahn MW, Demuth JP, Han SG. 2007. Accelerated rate of gene gain and loss in primates. Genetics. 177:1941–1949.

Hahn MW, Han MV, Han SG. 2007. Gene family evolution across 12 Drosophila genomes. PLoS Genet. 3:2135–2146.

Hardies SC, Edgell MH, Hutchison CA. 1984. Evolution of the mammalian β-globin gene cluster. J Biol Chem. 259: 3748–3756.

Hardison R. 1984. Comparison of the β-like globin gene families of rabbits and humans indicates that the gene cluster 5′ ε-γ-δ-β 3′ predates the mammalian radiation. Mol Biol Evol. 1:390–410.

Hardison R. 1991. Evolution of globin gene families. In: Selander RK, Whittam TS, Clark AG, editors. Evolution at the molecular level. Sunderland (MA): Sinauer Associates. p. 272–289.

Hardison R. 1998. Hemoglobins from bacteria to man: evolution of different patterns of gene expression. J Exp Biol. 201: 1099–1117.

Hardison R. 2001. Organization, evolution and regulation of the globin genes. In: Steinberg MH, Forget BG, Higgs DR, Nagel RL, editors. Disorders of hemoglobin: genetics, pathophysiology, and clinical management. Cambridge: Cambridge University Press.

Hardison R, Gelinas RE. 1986. Assignment of orthologous relationships among mammalian α-globin genes by examining flanking regions reveals a rapid rate of evolution. Mol Biol Evol. 3:243–261.

Hardison R, Margot JB. 1984. Rabbit globin pseudogene ψβ2 is a hybrid of δ- and β-globin sequences. Mol Biol Evol. 1:302–316.

Hardison R, Miller W. 1993. Use of long sequence alignments to study the evolution and regulation of mammalian globin gene clusters. Mol Biol Evol. 10:73–102.

Hardison R, Slightom JL, Gumucio DL, Goodman M, Stojanovic N, Miller W. 1997. Locus control regions of mammalian β-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. Gene. 205:73–94.

Hill A, Hardies SC, Phillips PC, Davis MG, Hutchison CA, Edgell MH. 1984. Two mouse early embryonic β-globin gene sequences: evolution of the nonadult β-globins. J Biol Chem. 259:3739–3747.

Hoffmann FG, Opazo JC, Storz JF. 2008. Rapid rates of lineage-specific gene duplication and deletion in the α-globin gene family. Mol Biol Evol. 25:591–602.

Holloway K, Lawson VE, Jeffreys AJ. 2006. Allelic recombination and de novo deletions in sperm in the human beta-globin gene region. Hum Mol Genet. 15:1099–1111.

Jeffreys AJ, Barrie P, Harris S, Fawcett D, Nugent Z, Boyd AC. 1982. Isolation and sequence analysis of a hybrid δ-globin pseudogene from the brown lemur. J Mol Biol. 156:487–503.

Jobb G, von Haeseler A, Strimmer K. 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol Biol. 4:18.

Koop BF, Goodman M. 1988. Evolutionary and developmental aspects of two hemoglobin β-chain genes (ε$^M$ and β$^M$) of opossum. Proc Natl Acad Sci USA. 85:3893–3897.

Koop BF, Siemieniak D, Slightom JL, Goodman M, Dunbar J, Wright PC, Simons EL. 1989. Tarsius δ- and β-globin genes: conversions, evolution, and systematics. J Biol Chem. 264: 68–79.

Lam KWG, Jeffreys AJ. 2007. Processes of de novo duplication of human alpha-globin genes. Proc Natl Acad Sci USA. 104: 10950–10955.

Lam K-W, Jeffreys A. 2006. Processes of copy-number change in human DNA: the dynamics of α-globin gene deletion. Proc Natl Acad Sci USA. 103:8921–8927.

Long M, Betran E, Thornton K, Wang W. 2003. The origin of new genes: glimpses from the young and old. Nat Rev Genet. 4:865–875.

Loukopoulos D, Kollia P. 2001. Worldwide distribution of β thalassemia. In: Steinberg MH, Forget BG, Higgs DR, Nagel RL, editors. Disorders of hemoglobin: genetics, pathophysiology, and clinical management. Cambridge: Cambridge University Press. p. 861–877.

Lupski JR, Stankiewicz P. 2005. Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. PLoS Genet. 1:627–633.

Margot JB, Demers GW, Hardison R. 1989. Complete nucleotide sequence of the rabbit β-like globin gene cluster: analysis of intergenic sequences and comparison with the human β-like globin gene cluster. J Mol Biol. 205:15–40.

Metzenberg AB, Wurzer G, Huisman THJ, Smithies O. 1991. Homology requirements for unequal crossing over in humans. Genetics. 128:143–161.

Moon AM, Ley TL. 1990. Conservation of the primary structure, organization, and function of the human and mouse β-globin locus-activating regions. Proc Natl Acad Sci USA. 87: 7693–7697.

Murphy WJ, Eizirik E, O' Brien SJ, et al. (11 co-authors). 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. Science. 294:2348–2351.

Nozawa M, Nei M. 2007. Evolutionary dynamics of olfactory receptor genes in Drosophila species. Proc Natl Acad Sci USA. 104:7122–7127.

Opazo JC, Hoffmann FG, Storz JF. 2008a. Genomic evidence for independent origins of β-like globin genes in monotremes and therian mammals. Proc Natl Acad Sci USA. 105:1590–1595.

Opazo JC, Hoffmann FG, Storz JF. 2008b. Differential loss of embryonic globin genes during the radiation of placental mammals. Proc Natl Acad Sci USA. 105:12950–12955.

Ovcharenko I, Loots GG, Giardine BM, Hou MM, Ma J, Hardison RC, Stubbs L, Miller W. 2005. Mulan: multiple-sequence local alignment and visualization for studying function and evolution. Genome Res. 15:184–194.

Paunesku T, Stevanovic M, Radosavljevic D, Drmanac R, Crkvenjakov R. 1990. Origin of rat β-globin haplotypes containing three and five genes. Mol Biol Evol. 7:407–422.

Prasad AB, Allard MW, NISC Comparative Sequencing Program, Green ED. 2008. Confirming the phylogeny of mammals by use of large comparative sequence datasets. Mol Biol Evol. 25:1795–1808.

Prychitko T, Johnson RM, Wildman DE, Gumucio DL, Goodman M. 2005. The phylogenetic history of new world monkey β-globin reveals a platyrrhine β to δ gene conversion in the atelid ancestry. Mol Phylogenet Evol. 35:225–234.

Rodriguez F, Oliver JL, Marin A, Medina JR. 1990. The general stochastic model of nucleotide substitution. J Theor Biol. 142:485–501.

Ronquist F, Huelsenbeck JP. 2003. MrBayes3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.

Schimenti JC, Duncan CH. 1985. Structure and organization of the bovine β-globin genes. Mol Biol Evol. 2:514–525.

Schneider JA, Peto TEA, Boone RA, Boyce AJ, Clegg JB. 2002. Direct measurement of the male recombination fraction in the human beta-globin hot spot. Hum Mol Genet. 11:207–215.

Schwartz S, Elnitski L, Li M, Weirauch M, Riemer C, Smit A, NCS Program, Green ED, Hardison RC, Miller W. 2003. MultiPipMaker and supporting tools: alignments and analysis of multiple genomic DNA sequences. Nucleic Acids Res. 31:3518–3524.

Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W. 2000. PipMaker—a Web server for aligning two genomic DNA sequences. Genome Res. 10:577–586.

Shaw CJ, Lupski JR. 2004. Implications of human genome architecture for rearrangement-based disorders: the genomic basis for disease. Hum Mol Genet. 13:R57–R64.

Shehee WR, Loeb DD, Adey NB, et al. (15 co-authors). 1989. Nucleotide sequence of the BALB/c mouse β-globin complex. J Mol Biol. 205:41–62.

Steppan SJ, Adkins RM, Anderson J. 2004. Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes. Syst Biol. 53:533–553.

Stevanovic M, Paunesku T, Radosavljevic D, Drmanac R, Crkvenjakov R. 1989. Variant chromosomal arrangement of adult β-globin genes in rat. Gene. 79:139–150.

Storz JF, Baze M, Waite JL, Hoffmann FG, Opazo JC, Hayes JP. 2007. Complex signatures of selection and gene conversion in the duplicated globin genes of house mice. Genetics. 177:481–500.

Storz JF, Hoffmann FG, Opazo JC, Moriyama H. 2008. Adaptive functional divergence among triplicated α-globin genes in rodents. Genetics. 178:1623–1638.

Storz JF, Sabatino SJ, Hoffmann FG, Gering EJ, Moriyama H, Ferrand N, Monteiro B, Nachman MW. 2007. The molecular basis of high-altitude adaptation in deer mice. PLoS Genet. 3:e45.

Tagle DA, Slightom JL, Jones RT, Goodman M. 1991. Concerted evolution led to high expression of a prosimian primate δ globin gene locus. J Biol Chem. 266:7469–7480.

Tatusova TA, Madden TL. 1999. BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. FEMS Microbiol Lett. 174:247–250.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876–4882.

Townes TM, Fitzgerald MC, Lingrel JB. 1984. Triplication of a four-gene set during evolution of the goat β-globin locus produced three genes now expressed differentially during development. Proc Natl Acad Sci USA. 81:6589–6593.

Wood WG. 2001. Hereditary persistance of fetal hemoglobin and δβ thalassemia. In: Steinberg MH, Forget BG, Higgs DR, Nagel RL, editors. Disorders of hemoglobin: genetics, pathophysiology, and clinical management. Cambridge: Cambridge University Press. p. 356–388.

Zhang J, Dyer KD, Rosenberg HF. 2000. Evolution of the rodent eosinophil-associated RNase gene family by rapid gene sorting and positive selection. Proc Natl Acad Sci USA. 97:4701–4706.