Repeated Evolution of Chimeric Fusion Genes in the β-Globin Gene Family of Laurasiatherian Mammals

Michael J. Gaudry[1], Jay F. Storz[2], Gary Tyler Butts[3], Kevin L. Campbell[1], and Federico G. Hoffmann[3,4,*]

[1]Department of Biological Sciences, University of Manitoba, Winnipeg, MB, Canada

[2]School of Biological Sciences, University of Nebraska, Lincoln

[3]Department of Biochemistry, Molecular Biology, Entomology, and Plant Pathology, Mississippi State University

[4]Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University

*Corresponding author: E-mail: federico.g.hoffmann@gmail.com.

Accepted: May 5, 2014

Abstract

The evolutionary fate of chimeric fusion genes may be strongly influenced by their recombinational mode of origin and the nature of functional divergence between the parental genes. In the β-globin gene family of placental mammals, the two postnatally expressed δ- and β-globin genes (*HBD* and *HBB*, respectively) have a propensity for recombinational exchange via gene conversion and unequal crossing-over. In the latter case, there are good reasons to expect differences in retention rates for the reciprocal *HBB*/*HBD* and *HBD*/*HBB* fusion genes due to thalassemia pathologies associated with the *HBD*/*HBB* "Lepore" deletion mutant in humans. Here, we report a comparative genomic analysis of the mammalian β-globin gene cluster, which revealed that chimeric *HBB*/*HBD* fusion genes originated independently in four separate lineages of laurasiatherian mammals: Eulipotyphlans (shrews, moles, and hedgehogs), carnivores, microchiropteran bats, and cetaceans. In cases where an independently derived "anti-Lepore" duplication mutant has become fixed, the parental *HBD* and/or *HBB* genes have typically been inactivated or deleted, so that the newly created *HBB*/*HBD* fusion gene is primarily responsible for synthesizing the β-type subunits of adult and fetal hemoglobin (Hb). Contrary to conventional wisdom that the *HBD* gene is a vestigial relict that is typically inactivated or expressed at negligible levels, we show that *HBD*-like genes often encode a substantial fraction (20–100%) of β-chain Hbs in laurasiatherian taxa. Our results indicate that the ascendancy or resuscitation of genes with *HBD*-like coding sequence requires the secondary acquisition of *HBB*-like promoter sequence via unequal crossing-over or interparalog gene conversion.

Key words: β-globin, concerted evolution, gene conversion, gene duplication, gene family evolution, hemoglobin, Laurasiatheria.

Introduction

The probability that chimeric fusion genes are retained in the genome may be strongly influenced by their recombinational mode of origin and the nature of functional divergence between the parental genes (Katju and Lynch 2003, 2006; Jones and Begun 2005; Jones et al. 2005; Rogers et al. 2009, 2010; Kaessmann 2010; Katju 2012, 2013). Unequal crossing-over (nonallelic homologous recombination) between tandem gene duplicates represents a common mechanism for producing chimeric fusion genes in conjunction with changes in gene copy number (Holloway et al. 2006; Hoffmann et al. 2008b). In cases where the breakpoint of an unequal cross-over occurs at homologous sites in a misaligned pair of tandem gene duplicates, both recombinant chromosomes will contain chimeric genes with reciprocal fusions of paralogous sequence. One recombinant chromosome (the duplication mutant) will harbor a unique chimeric fusion gene flanked by intact copies of the parental genes on either side, whereas the other recombinant chromosome (the deletion mutant) will harbor a solitary chimeric gene with the reciprocal fusion of coding sequence from each of the two parental genes (fig. 1*A*). If the differences in gene content between the two recombinant chromosomes affect fitness, then the deletion mutant and duplication mutant will have different probabilities of evolutionary persistence. Similarly, variation in functional constraint may explain patterns of differential retention among the three genes on the duplication

chromosome. If members of the parental gene pair are differentially expressed—due to differences in proximity to a distal cis-regulatory element and/or differences in proximal cis-regulatory sequence—then the newly created fusion gene and the two repositioned copies of the parental genes may have distinct expression profiles at their inception and may thus have different probabilities of loss or fixation. If members of the parental gene pair have different coding sequences, then the nascent paralogs on the duplication chromosome will be structurally distinct at their inception, which may influence their probabilities of loss or fixation. Chimeric fusion genes that incorporated distinct functional modules of two separate parental genes are known to have evolved novel functions in a diverse range of organisms (reviewed by Long et al. 2003; Fan et al. 2008; Hahn 2009; Kaessmann 2010; Corduso-Moreira and Long 2012; Hoogewijs et al. 2012; Katju 2012).

The β-globin gene family of placental mammals provides an excellent system for investigating how the evolutionary fates of chimeric fusion genes may be influenced by their recombinational mode of origin. The β-globin gene cluster of placental mammals contains a set of developmentally regulated genes that are arranged in their temporal order of expression and typically include three genes at the 5′-end of the cluster, ε-globin (*HBE*), γ-globin (*HBG*), and η-globin (*HBH*), which are expressed in embryonic and/or fetal erythroid cells, and two genes at the 3′-end of the cluster, δ-globin (*HBD*) and β-globin (*HBB*), which are expressed in adult and fetal erythroid cells (Hardison 2001, 2012). Interspecific variation in the size and membership composition of the β-globin gene family is attributable to lineage-specific gene gains via duplication and lineage-specific gene losses via deletion or inactivation (Hoffmann et al. 2008b; Opazo et al. 2008a, 2008b; Storz et al. 2011, 2013; Hardison 2012). The *HBD* and *HBB* paralogs represent the products of a tandem gene duplication that occurred in the stem lineage of placental mammals (Goodman et al. 1984; Hardison 1984; Opazo et al. 2008a, 2008b; Hoffmann et al. 2010). In humans and other mammals that have been investigated to date, *HBB* is typically expressed at a much higher level than *HBD* because it is under the transcriptional control of a stronger basal promoter (Poncz et al. 1983; Antoniou and Grosveld 1990; Hardison 2001). Moreover, the *HBD* mRNA has a shorter half-life than that of *HBB* (Forget 2001). Thus, in most species that have retained intact copies of both *HBD* and *HBB*, the β-type subunits of postnatally expressed hemoglobin (Hb) are primarily encoded by one or more copies of the *HBB* gene.

HBD and *HBB* have a propensity for recombinational exchange via gene conversion and unequal crossing-over (fig. 1), and these exchanges appear to be highly asymmetric, as the coding region of *HBD* has been converted by the downstream *HBB* gene in multiple lineages, particularly in the 5′-coding region (Jeffreys et al. 1982; Martin et al. 1983; Goodman et al. 1984; Hardies et al. 1984; Hardison 1984; Hardison and Margot 1984; Koop et al. 1989; Tagle et al. 1991; Prychitko et al. 2005; Hoffmann et al. 2008a; Opazo et al. 2008a). However, these events typically result in *HBD* coding sequence that is fused to *HBB*-like upstream sequence that does not extend to the *HBB* CCAAT promoter element (Hardies et al. 1984; Koop et al. 1989), so that expression levels of the resultant fusion gene are not altered. Despite only a few known examples (e.g., paenungulates; Opazo et al. 2009), there are good reasons to expect asymmetry in

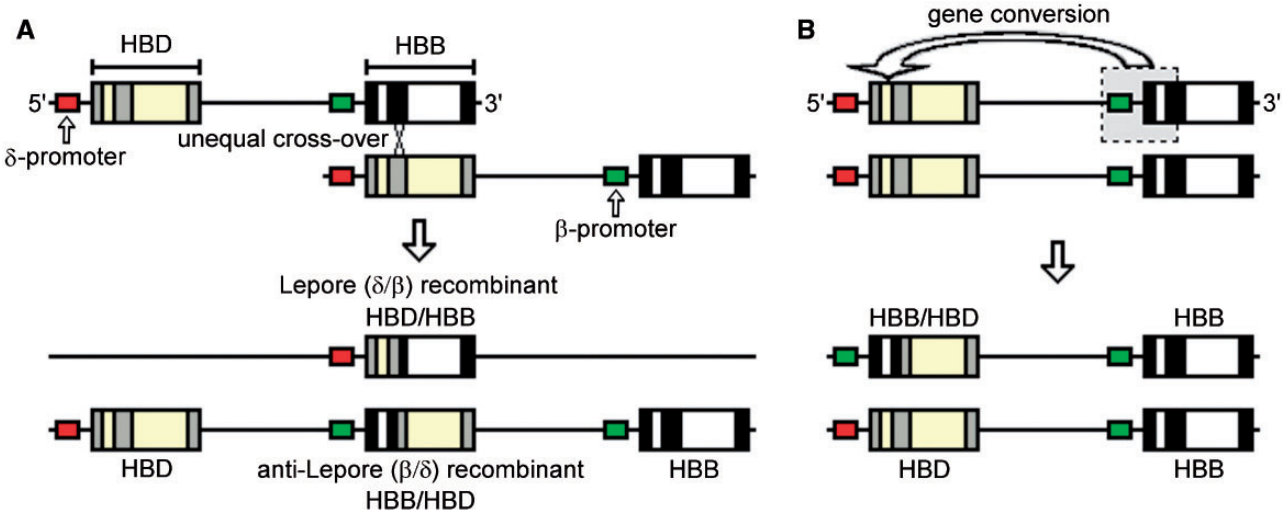

Fig. 1.—Chimeric fusion genes in the mammalian β-globin gene cluster can be produced via two separate recombinational mechanisms. (*A*) Unequal crossing-over between a misaligned pair of *HBD* and *HBB* paralogs can produce Lepore and anti-Lepore recombinant chromosomes. (*B*) Interparalog gene conversion between *HBD* and *HBB* can also produce chimeric fusion genes that are structurally similar to the Lepore and anti-Lepore fusion genes but without the associated changes in gene copy number.

the fixation or retention of chimeric fusion genes that result from crossovers between misaligned copies of *HBD* and *HBB*. In humans, the products of these rare crossovers result in a solitary *HBD/HBB* fusion gene on one recombinant chromosome (the Hb Lepore deletion mutant) and the reciprocal *HBB/HBD* fusion gene on the other recombinant chromosome (the anti-Lepore duplication mutant; Forget 2001). In the former case, the *HBD/HBB* fusion gene is solely responsible for synthesizing the β-type subunits of adult Hb, and in the latter case, the reciprocal *HBB/HBD* fusion gene is flanked by functionally intact copies of the parental *HBD* gene on the 5′-side and the parental *HBB* gene on the 3′-side (fig. 1*A*). Heterozygous carriers of the Hb Lepore mutation produce red blood cells that contain normal $\alpha_2\beta_2$ Hb tetramers in addition to lesser quantities of $\alpha_2(\delta/\beta)_2$ tetramers that incorporate β-chain products of the chimeric fusion gene. The lower abundance of the Hb Lepore isoform is mainly due to the lower transcription rate of the *HBD/HBB* fusion gene, as it is under the control of the weak *HBD* promoter (Forget 2001). Hb Lepore heterozygotes suffer from a mild form of hemolytic anemia, whereas homozygotes suffer from far more serious forms of erythrocytic dysfunction caused by an imbalance of α- and β-chain synthesis (Olivieri and Weatherall 2001). The dosage imbalance results in insoluble aggregations of oxidized α-chain monomers and their cytotoxic breakdown products (iron, heme, and hemichrome) in erythroid precursor cells and mature erythrocytes, which leads to premature hemolysis (Rachmilewitz and Schrier 2001). In contrast to the "δ/β-thalassemia" disease phenotype associated with the Hb Lepore deletion mutant, inheritance of the anti-Lepore duplication is not associated with any hematological pathology (Wood 2001; Saller et al. 2012). The well-documented fitness consequences of the Hb Lepore deletion mutation in humans suggest that independently derived *HBB/HBD* and *HBD/HBB* fusion genes in other mammalian species can be expected to have different fixation probabilities.

Here, we report the results of a comparative genomic analysis of the mammalian β-globin gene cluster that sheds light on the origins and phylogenetic distribution of chimeric fusion genes. First, our analysis revealed that functional *HBB/HBD* fusion genes have originated via unequal crossing-over at least three times independently in the Laurasiatheria, a supraordinal clade that contains Chiroptera, Eulipotyphla, Pholidota, Carnivora, Perissodactyla, and Cetartiodactyla. In contrast, functionally intact copies of the reciprocal *HBD/HBB* fusion gene appear to be completely absent. Second, contrary to conventional wisdom that the *HBD* gene is a vestigial relict that is always either inactivated or expressed at negligible levels (Martin et al. 1983; Hardies et al. 1984; Hardison 1984; Hardison and Margot 1984; Koop et al. 1989; Prychitko et al. 2005), we identified a surprisingly large number of laurasiatherian taxa in which the β-type subunits of adult-expressed Hb contain *HBD*-like primary structures. Taken together, our results confirm that the retention and ascendancy of genes with *HBD*-like coding sequence require the retention of *HBB*-like promoter sequence via unequal crossing-over or the secondary acquisition of *HBB*-like promoter sequence via gene conversion.

Materials and Methods

Annotation of Genomic Sequences

We obtained genomic sequences containing the β-globin gene cluster from 35 species representing each of the major lineages of laurasiatherian mammals. All sequences were obtained from GenBank and Ensembl. A list of all examined laurasiatherian species and the accession numbers for all associated sequences are provided in supplementary table S1, Supplementary Material online.

In the genome assembly of each species, we identified β-like globin genes in unannotated sequences by using the program Genscan (Burge and Karlin 1997) and by comparing known exon sequences with genomic contigs using the program Blast2 version 2.2 (Tatusova and Madden 1999). Globin-like open-reading frames were considered to be putatively functional if they had conserved exon length and conserved splice sites and if they lacked premature stop codons and frame-shift mutations. Genes were classified according to their similarity to genes in the human globin gene clusters, which were used as reference standards for all comparisons.

Inferring Orthologous Relationships and Identifying Cases of Interparalog Gene Conversion

To assign orthologous relationships among genes and specific gene regions, and to identify cases of interparalog gene conversion, we conducted pairwise comparisons of sequence similarity with the human gene cluster using the program Pipmaker (Schwartz et al. 2000, 2003). We also conducted separate phylogenetic reconstructions based on coding sequence, intron 2 sequence, and the 5′ and 3′-flanking sequences of each gene. Specifically, we conducted phylogenetic analyses on four discrete partitions of a multiple sequence alignment: 500 bp of 5′-flanking sequence immediately upstream of the initiation codon, 444 bp of coding sequence, the complete intron 2 sequence (which varies in length among the different genes included), and 500 bp of 3′-flanking sequence immediately downstream of the termination codon. Analyses based on noncoding sequence were restricted to the *HBB* and *HBD* genes. The rationale for conducting phylogenetic analyses on each of these different data partitions is that interparalog gene conversion between tandemly duplicated globin genes is typically restricted to coding sequence, so noncoding flanking sequence typically records the most accurate history of gene duplication and species divergence (Hardison and Gelinas 1986; Hoffmann and Storz 2007; Storz et al. 2007, 2008, 2009, 2010, 2012; Hoffmann et al. 2008a, 2008b; Opazo et al. 2008a, 2008b, 2009; Runck

et al. 2009, 2010). In the case of mammalian *HBD*, ectopic recombinational exchanges are typically restricted to the 5′-end of the gene as *HBB* → *HBD* conversion events typically overwrite exon 1, intron 1, and exon 2 of the *HBD* recipient sequence (Hardies et al. 1984; Hardison 1984; Hardison and Margot 1984; Prychtiko et al. 2005; Hoffmann et al. 2008a; Opazo et al. 2008b, 2009). Thus, sequence variation in intron 2 and the 3′-flanking region is best suited to the task of assigning orthology, and comparisons of 5′- and 3′-flanking regions can reveal whether chimeric fusion genes were created via unequal crossing-over or gene conversion (Hoffmann et al. 2008b; Opazo et al. 2009). With the exception of chimeric fusion genes, we classified each gene as being *HBB*-like or *HBD*-like on the basis of intron 2 sequences using human *HBD* as a reference standard.

Phylogenetic analyses for all the different partitions were performed according to the following bioinformatic protocol. Sequences were aligned using the L-INS-i strategy from Mafft v7 (Katoh and Standley 2013). We performed maximum-likelihood analyses in Treefinder, version March 2011 (Jobb et al. 2004), evaluating support for the nodes with 1,000 bootstrap pseudoreplicates. We used the "propose model" tool of Treefinder to select the best-fitting models of nucleotide substitution based on the Akaike information criterion with correction for small sample size. We estimated Bayesian phylogenies in Mr. Bayes v. 3.1.2 (Ronquist and Huelsenbeck 2003), running six simultaneous chains for 2×107 generations, sampling every 2.5×103 generations, and using default priors. A given run was considered to have reached convergence once the likelihood scores reached an asymptotic value and the average standard deviation of split frequencies remained <0.01. We discarded all trees that were sampled prior to convergence, and we evaluated support for the nodes and parameter estimates from a majority rule consensus of the last 2,500 trees.

Conserved *cis*-regulatory elements (distal and proximal CA CCC, CCAAT, and TATA boxes) that are known to be essential for high-level expression of β-like globin genes (Myers et al. 1986; Ebb et al. 1998; Ristaldi et al. 1999) were manually annotated for *HBD*- and *HBB*-like genes in regions 150 bp upstream of the putative Cap sites (typically located ~50 bp upstream of the initiation codon).

Our inferences of orthology and paralogy were refined by comparing phylogenetic reconstructions with the context and content orthology inferences based on CHAP2 (Song et al. 2012). The CHAP2 analyses were based on the phylogeny from Meredith et al. (2011) and were restricted to a subset of laurasiatherian species for which we had complete or mostly complete sequence coverage of the β-globin gene cluster. Using an alternative tree topology congruent with that proposed by Nery et al. (2012) yielded similar results. Complete results for the CHAP2 analyses are available upon request.

Results

Patterns of Gene Turnover in the Eutherian β-Globin Gene Cluster

We obtained genomic sequences corresponding to the β-globin gene cluster from 75 placental mammals representing each of the four supraordinal clades: Afrotheria (7 species), Xenarthra (2 species), Euarchontoglires (14 glires + 17 primates), and Laurasiatheria (35 species). The initial survey revealed a preponderance of chimeric fusion genes in laurasiatherian taxa, so this group served as the main focus for all subsequent analyses. Comparison of the β-globin gene clusters among the laurasiatherian species in our study revealed considerable variation in gene copy number (fig. 2). The number of pseudogenes was variable as well, ranging from 0 in most bats to 7 in the goat (*Capra aegagrus hircus*). Consistent with previous surveys based on smaller numbers of mammalian taxa (Hoffmann et al. 2008b; Opazo et al. 2008a, 2008b, 2009), the 5′-end of the cluster contains the prenatally expressed genes, *HBE*, *HBG*, and *HBH*; all species examined possess one or two copies of *HBE* located upstream of one or more copies of an additional embryonic gene—either *HBG* or *HBH*. Similarly, the 3′-end of the cluster contains the postnatally expressed genes, *HBD* and *HBB*. The β-globin gene cluster of bovids represents the only exception to this general pattern, as one or more en bloc duplications have transposed some early-expressed *HBE* and/or *HBH* genes to chromosomal locations upstream of one or more late-expressed *HBB* genes (Townes et al. 1984; Schimenti and Duncan 1985; fig. 2).

In contrast to taxa in the Euarchontoglires clade (Primates, Rodentia, Lagomorpha, Dermoptera, and Scandentia), which typically possess 1–2 functional copies of *HBG* and, in some taxa, a single *HBH* pseudogene, most laurasiatherians possess functional copies of *HBH* as an additional early-expressed gene, whereas a few lineages have retained a single *HBG* pseudogene (fig. 2). In the case of the late-expressed (adult) genes at the 3′-end of the gene cluster, most mammalian species possess one or two copies of *HBD* and/or *HBB*. The *HBD* gene has been independently inactivated (but never deleted) in many species of Euarchontoglires, whereas functionally intact copies seem to have been retained in the majority of laurasiatherian species examined (fig. 2). Thus, most variation in the size and membership composition of the β-globin gene family is attributable to the differential loss or inactivation of the embryonic *HBG* and *HBH* genes and the late-expressed *HBD* gene.

Phylogenetic analyses based on coding sequence enabled us to resolve orthology for the early-expressed genes of laurasiatherians, with *HBE*, *HBG*, and *HBH* sequences clustering into reciprocally monophyletic groups (fig. 3). However, these analyses could not resolve orthologous relationships for the late-expressed *HBD* and *HBB* genes, as paralogs from the same species typically showed higher levels of sequence

similarity with one another than with positional orthologs in other species (fig. 4*A*). This phylogenetic pattern is consistent with a complex history of lineage-specific gene turnover and interparalog gene conversion. To gain more refined insights into the evolutionary history of *HBD* and *HBB*, and to identify chimeric sequences that result from recombinational exchanges between the two paralogs that extend beyond the exons of these genes, we compared phylogenetic trees estimated from coding sequence with trees estimated from three discrete partitions of noncoding sequence: 500 bp of upstream flanking sequence, intron 2, and 500 bp of downstream flanking sequence (fig. 4*B–D*). These analyses included a number of truncated genes in addition to those with fully intact reading frames. With the exception of a single pseudogene (*HBB*-T4 ps) in the bovine gene cluster, analyses of intron 2 sequences reliably grouped all sequences with either the human *HBD* or *HBB* gene (fig. 4*C*).

Patterns of Chimerism

For each gene, the 5′-flanking sequence, intron 2, and 3′-flanking sequence were classified according to similarity with homologous sequence in the human *HBB* or *HBD* genes. Out of 70 examined genes, 40 genes exhibited clear affinities to human *HBD* or *HBB* in each of the three noncoding segments. In the remaining 30 cases, the different

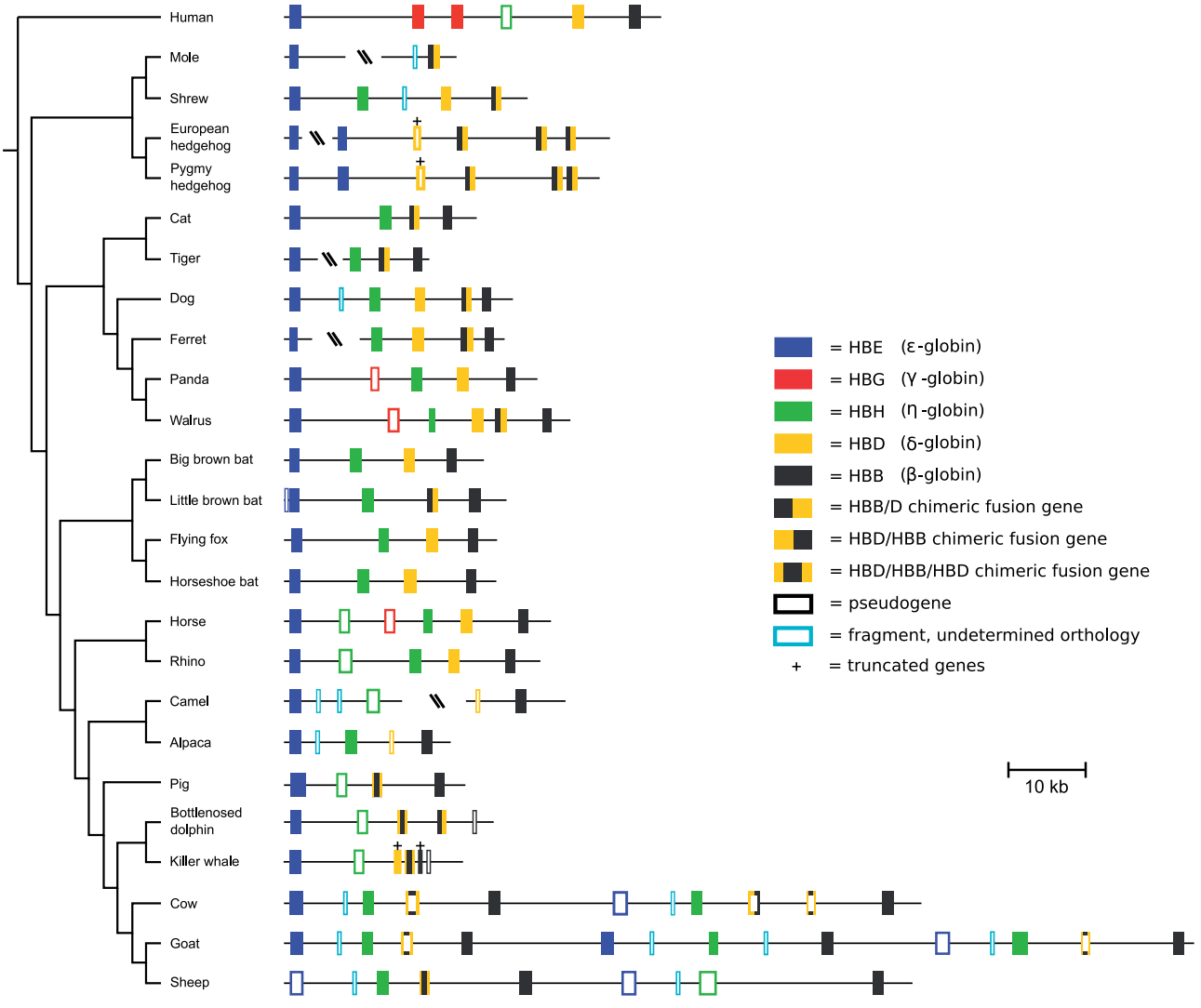

Fig. 2.—Genomic structure of the β-globin gene cluster in 24 laurasiatherian mammals, with the orthologous gene cluster from human provided as an outgroup. Each of the major laurasiatherian lineages are represented, including eulipotyphlans (moles, shrews, and hedgehogs), carnivores, bats (including microchiroptera, megachiroptera, and yingchiroptera), perissodactyls, and cetartiodactyls. Species were not included if the genome assemblies lacked sufficient coverage to determine the linkage order of genes in the β-globin gene cluster. Paired forward slashes denote sequence coverage gaps. The tree topology is based on Meredith et al. (2011).

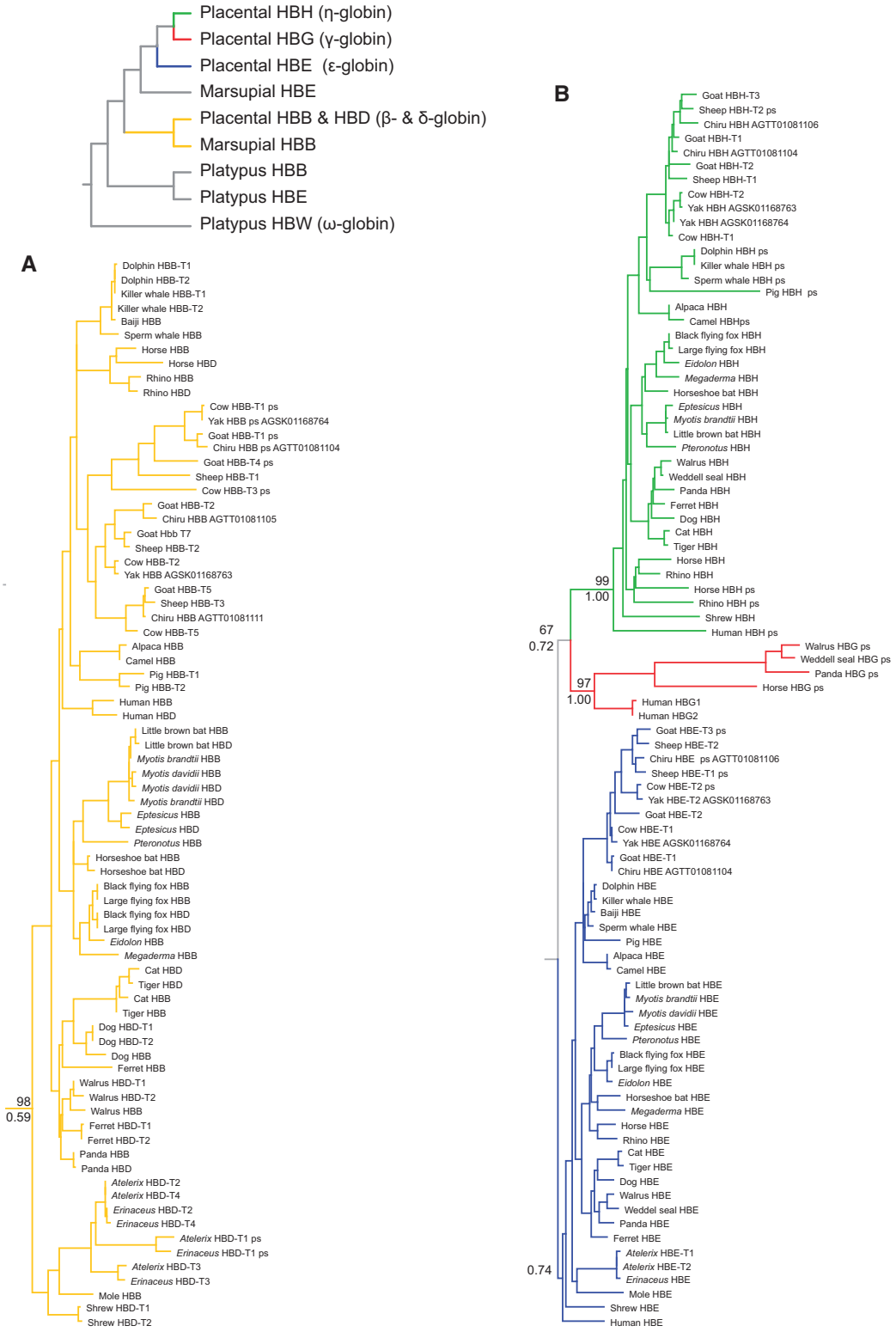

Fig. 3.—Maximum-likelihood phylogeny depicting relationships among the β-like globins genes of laurasiatherian mammals based on an alignment of coding sequences. Repertoires of β-like globin genes from human, gray short-tailed opossum (*Monodelphis domestica*), and platypus (*Ornithorhynchus anatinus*) were included for comparison. Pseudogenes are indicated by the abbreviation "ps." Maximum-likelihood bootstrap support (above) and Bayesian posterior probabilities (below) are provided next to the relevant nodes. (*A*) The subtree of the β- and δ-like globins and (*B*) the subtree of ε-, γ-, and η-like globins. The inset on top shows the phylogeny of mammalian β-like globins.

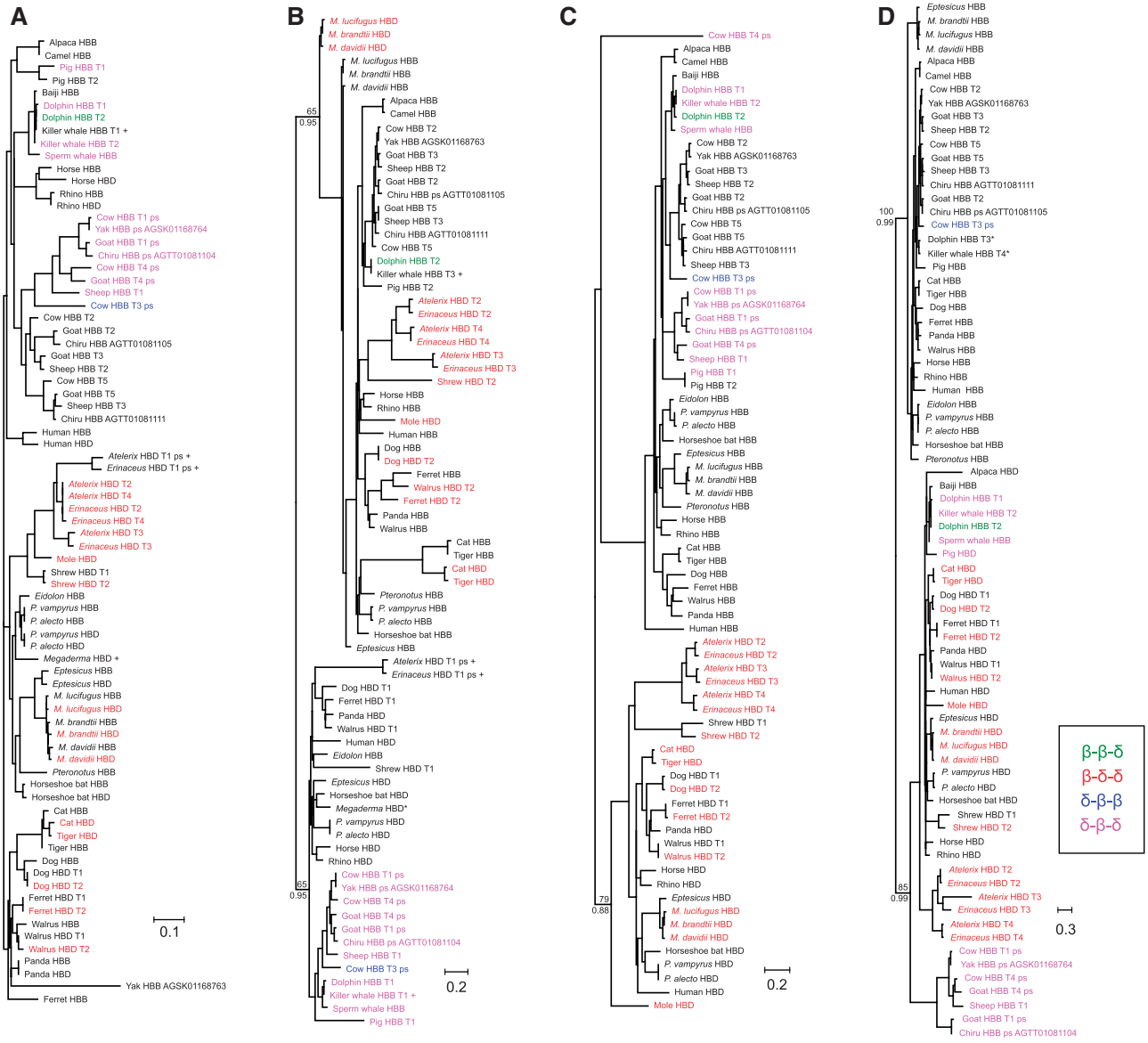

Fig. 4.—Maximum-likelihood phylogenies depicting relationships among *HBD* and *HBB* genes based on coding sequence (*A*), 500 bp of 5′-flanking sequence (*B*), intron 2 sequence (*C*), and 500 bp of 3′-flanking sequence (*D*). Because the β-like globin genes have undergone multiple rounds of duplication that have resulted in tandemly repeated sets of paralogous gene copies, we index each paralog with the notation "TX," where X is a number corresponding to the linkage order of each gene in the 5′- to 3′-orientation (Aguileta et al. 2006); pseudogenes denoted by the abbreviation ps. Maximum-likelihood bootstrap support (above) and Bayesian posterior probabilities (below) are provided next to the relevant nodes.

noncoding segments of the same gene were not congruent in their affinities for human *HBD* or *HBB*, indicating that they are chimeric fusion genes. In this set of 30 chimeric genes, we observed four of the six possible chimeric combinations of *HBD*-like and *HBB*-like noncoding segments (table 1). Although unequal crossing-over should produce equal numbers of *HBB*/*HBD* (anti-Lepore) and *HBD*/*HBB* (Lepore) fusion genes, examination of sequence variation in three noncoding segments (5′-flanking sequence, intron 2, and 3′-flanking sequence) revealed a disproportionate number of functionally intact fusion genes with *HBB*-like 5′-flanking sequences relative to those with *HBD*-like 5′-flanking sequence (table 1 and supplementary table S2, Supplementary Material online). This pattern suggests that *HBB*/*HBD* fusion genes are less dispensable than the reciprocal *HBD*/*HBB* fusion genes, perhaps because *HBB*-like promoter sequence is required for high-level expression. There are additional observations consistent with this hypothesis: 1) In all examined species, late-expressed β-like globin genes have retained an *HBB*-like upstream sequence, 2) *HBB*-like genes with mutations in upstream

Table 1
Patterns of Sequence Chimerism in a Sample of 70 Late-Expressed β-Like Globin Genes in Laurasiatherian Mammals

Cross-Over Type	Chimeric Pattern (5′-Intron 2-3′)	Genes	Pseudogenes	
	β-β-β	30	—	
Anti-Lepore	β-β-δ	1	—	Cetacea (*Tursiops*)
	β-δ-β	—	—	
	β-δ-δ	16	—	Carnivora, Chiroptera (*Myotis*), Eulipotyphla
Lepore	δ-β-β	—	1	Cetartiodactyla (Cow)
	δ-β-δ	4	7	Cetartiodactyla
	δ-δ-β	—		
	δ-δ-δ	11		

NOTE.—The classification of chimeric patterns is based on sequence matches between noncoding segments (5′-flanking sequence, intron 2, and 3′-flanking sequence) and the homologous segments of the human *HBD* and *HBB* genes (see text for details). The reciprocal *HBD*/*HBB* ("*δ-β-β*" and "*δ-δ-β*") and *HBB*/*HBD* ("*β-β-δ*" and "*β-δ-δ*") fusion genes are described as possible products of "Lepore" and "anti-Lepore" crossovers (fig. 1), but in any given case, the same pattern of sequence chimerism could have been produced by *HBB* → *HBD* or *HBD* → *HBB* gene conversion. The nonchimeric "β-β-β" and "δ-δ-δ" genes represent cases where each of the three noncoding segments match the corresponding segments of the human *HBB* and *HBD* genes, respectively.

regulatory elements (e.g., distal CACCC and TATA boxes of ferret *HBB*) do not appear to be transcribed/translated, and 3) over half of the chimeric fusion genes with *HBD*-like upstream sequence are pseudogenes. There do not appear to be any cases where the major adult Hb isoform is encoded by a gene with *HBD*-like upstream sequence (fig. 5).

Independent Origins of Chimeric HBB/HBD Fusion Genes

The *HBB*/*HBD* fusion genes of the bottlenose dolphin (*Tursiops truncatus*), eulipotyphlans, and carnivores appear to represent "anti-Lepore" duplicates, where the 5′-sequence derives from an *HBB*-like gene, and the 3′-sequence derives from an *HBD*-like gene. All of these *HBB*/*HBD* fusion genes have intact reading frames. Each of these identified *HBB*/*HBD* fusion genes have upstream flanking sequences that are *HBB*-like, and intron 2 and downstream sequences that are *HBD*-like, with the sole exception of the dolphin fusion gene, which—for reasons explained below—has an *HBB*-like intron 2 sequence (fig. 4*C*, table 1, and supplementary table S2, Supplementary Material online).

The identified *HBB*/*HBD* fusion genes are equally similar to the human *HBD* and *HBB* genes at the 5′-end (exons 1 and 2 and intron 1), but they exhibit a higher sequence similarity with human *HBD* at the 3′-end (intron 2 and exon 3; fig. 6). It thus appears that pure, unadulterated *HBD* genes have not been retained in the β-globin gene clusters of any extant mammal, probably due to a long history of recurrent *HBB* → *HBD* gene conversion that may have occurred prior to some of the early branching events in the radiation of eutherian mammals. We can use intron 2 sequences and noncoding flanking sequences to identify true orthologs of human *HBD*, with the caveat that all such genes may be equally "*HBB*-like" and "*HBD*-like" in exons 1, 2, and 3 and intron 1.

In addition to chimeric fusion genes that originated via unequal crossing over, the *HBB*/*HBD* fusion gene in the microchiropteran bat genus *Myotis* appears to have originated via *HBB* → *HBD* gene conversion that extended approximately 240 bp upstream of the initiation codon (data not shown). Similarly, in the stem lineage of cetartiodactyls, a *HBB* → *HBD* gene conversion event occurred that spanned intron 2. Consequently, this fusion gene has *HBD*-like 5′- and 3′-flanking sequence in combination with *HBB*-like coding sequence and intron 2 sequence ("δ-β-δ" in table 1). Functional copies of this gene have been retained in the bottlenose dolphin, killer whale (*Orcinus orca*), sperm whale (*Physeter macrocephalus*), and pig (*Sus scrofa domesticus*); it became pseudogenized in tylopods and bovids. When an unequal cross-over later occurred in the common ancestor of cetaceans, the duplicated *HBB*/*HBD* gene on the anti-Lepore chromosome formed via fusion of 5′-*HBB* coding sequence to the 3′-end of the *HBD* gene whose intron 2 had previously been converted by *HBB*. This two-step process of *HBB* → *HBD* gene conversion followed by an anti-Lepore chimeric duplication appears sufficient to explain the mosaic sequence of the dolphin *HBB*/*HBD* fusion gene.

To validate inferences derived from the phylogenetic analysis, we used the CHAP2 package (Song et al. 2012) to make independent orthology assignments and to identify cases of interparalog gene conversion. Briefly, CHAP2 makes inferences of "context orthology" without accounting for the possibility of interparalog gene conversion, whereas "content orthology" tracks the history of each nucleotide within the alignment and also considers gene conversion events. These analyses were restricted to a subset of 21 species for which sequence coverage included a substantial portion of the β-globin gene cluster and used the human gene cluster as reference (supplementary table S1, Supplementary Material online). CHAP2 results were highly congruent with phylogeny-based inferences, with a few exceptions. For example, the gene just downstream from the *HBB*/*HBD* fusion gene in the cat β-globin cluster was identified as an *HBE* ortholog by CHAP2, whereas our phylogenetic analyses placed it within the *HBH* clade (fig. 7 and supplementary fig. S1, Supplementary Material online). In addition, the content orthology results from CHAP2 identify more fine-grained patchworks of mosaic sequence, which enabled us to detect additional cases of interparalog gene conversion. Results of this analysis identified ectopic conversion tracts in the *HBD* gene of panda (*Ailuropoda melanoleuca*), horse (*Equus caballus*), horseshoe bat (*Rhinolophus ferrumequinum*), and in the *HBB* gene of white rhinoceros (*Ceratotherium simum*), flying fox (*Pteropus vampyrum*), and big brown bat (*Eptesicus*

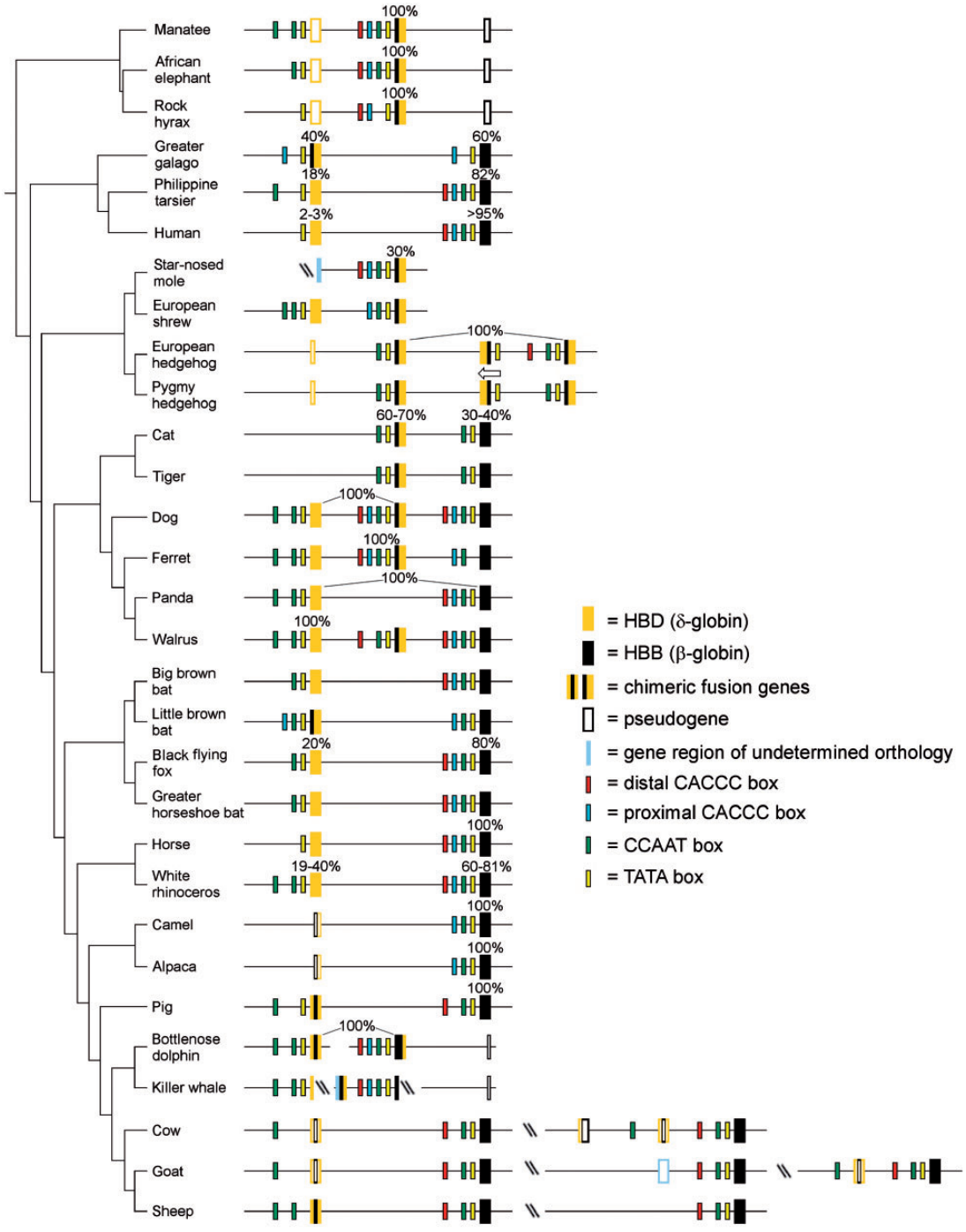

Fig. 5.—Annotation of cis-regulatory elements associated with *HBD*- and *HBB*-like genes in each of the clades of placental mammals in which functional chimeric *HBB*/*HBD* fusion genes have been identified: Afrotheria (including the paenungulates with anti-Lepore *HBB*/*HBD* fusion genes; Opazo et al. 2009), Primates (including the greater galago, *Otolemur crassicaudatus*, which possesses an *HBB*/*HBD* fusion gene that was produced via *HBB* → *HBD* gene conversion; Tagle et al. 1991), and the Laurasiatheria. Relative expression levels of alternative β-chain Hb isoforms were taken from the literature (supplementary table S3, Supplementary Material online). Each gene was classified based on phylogeny reconstructions of noncoding sequences (500 bp upstream, intron 2, and 500 bp downstream) shown in figure 4. Paired forward slashes denote sequence coverage gaps. The tree topology follows Meredith et al. (2011).

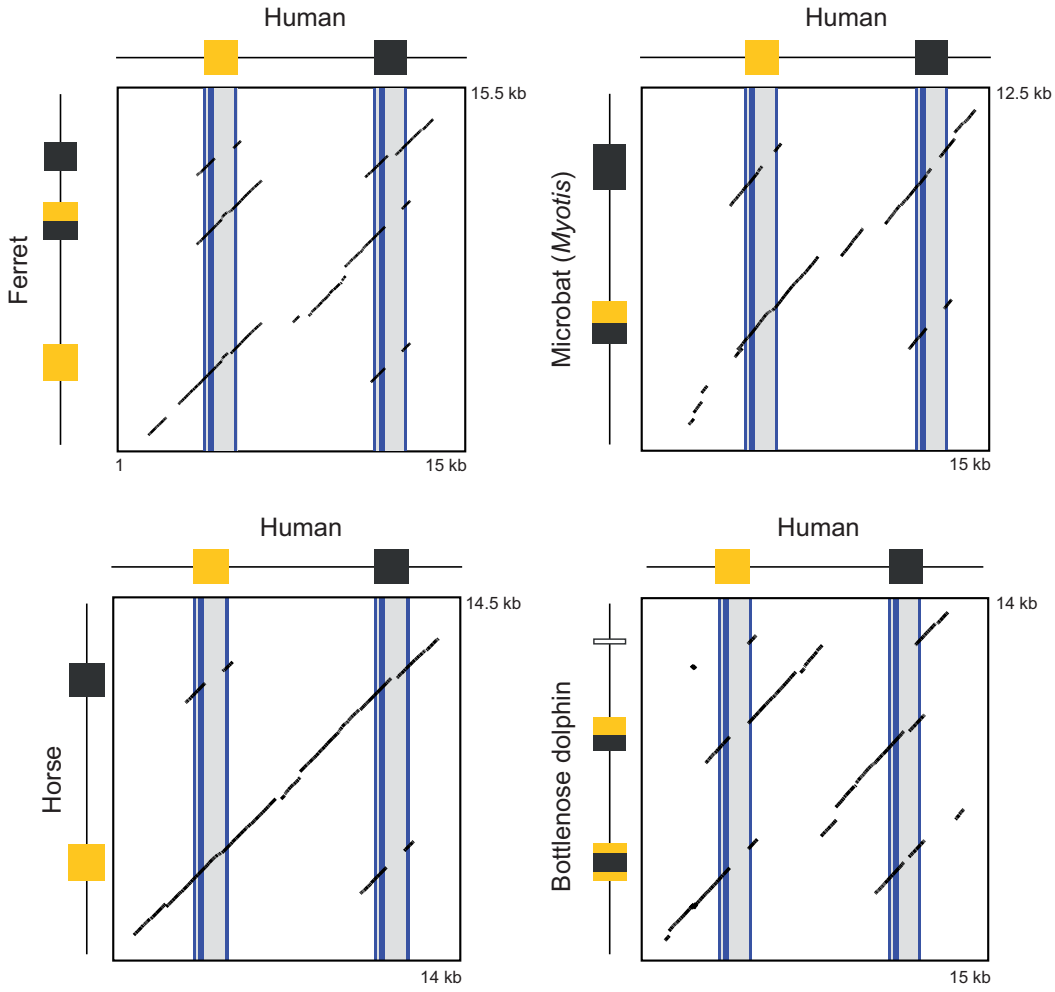

Fig. 6.—Dot plots of sequence similarity between the *HBD* and *HBB* genes of select laurasiatherian mammals and human. Top left: Ferret (*Mustela putorius furo*) genes versus human genes; top right: Little brown bat (*Myotis lucifugous*) genes versus human genes. Bottom left: Horse (*Equus ferus caballus*) genes versus human genes; bottom right: Bottlenose dolphin (*Tursiops truncatus*) genes versus human genes.

fuscus; fig. 7). Assignments based on context orthology are shown in supplementary fig. S1, Supplementary Material online.

Patterns of Gene Loss Following the Formation of Chimeric Fusion Genes

Following the duplicative origins of "anti-Lepore" *HBB*/*HBD* fusion genes, the parental *HBD* and *HBB* genes show a consistent pattern of inactivation/loss. Previous studies have documented that paenungulates (elephants, sea cows, and hyraxes) and eulipotyphlans have β-type Hb subunits that are exclusively encoded by *HBB*/*HBD* or *HBD*-like genes, respectively (Opazo et al. 2008b, 2009; Campbell et al. 2010, 2012; Signore et al. 2012). Paenungulates have a chimeric *HBB*/*HBD* fusion gene that is flanked by an *HBD* pseudogene on the 5′-side and an *HBB* pseudogene on the 3′-side, a rearrangement that is structurally similar to the anti-Lepore duplication mutation in humans. However, in paenungulates, the duplicated *HBB*/*HBD* fusion gene supplanted each of the parental gene copies and is therefore solely responsible for synthesizing the β-type subunits of adult and fetal Hb (Opazo et al. 2009). Here, we show that all *HBD*-like genes of eulipotyphlans have *HBB*-like upstream flanking sequence with the sole exception of the *HBD*-T1 gene in the Eurasian shrew (*Sorex araneus*). Available evidence thus suggests that the parental *HBB* gene was deleted soon after the formation of the chimeric *HBB*/*HBD* fusion gene in eulipotyphlans. The parental *HBD* gene was also inactivated in the ancestor of erinaceids (hedgehogs), followed by a reduplication of the chimeric *HBB*/*HBD* fusion gene (fig. 2). Similarly, the parental *HBD* gene was deleted in felids and the parental *HBB* gene was inactivated in toothed whales. Thus, in the majority of cases where anti-Lepore duplication chromosomes have

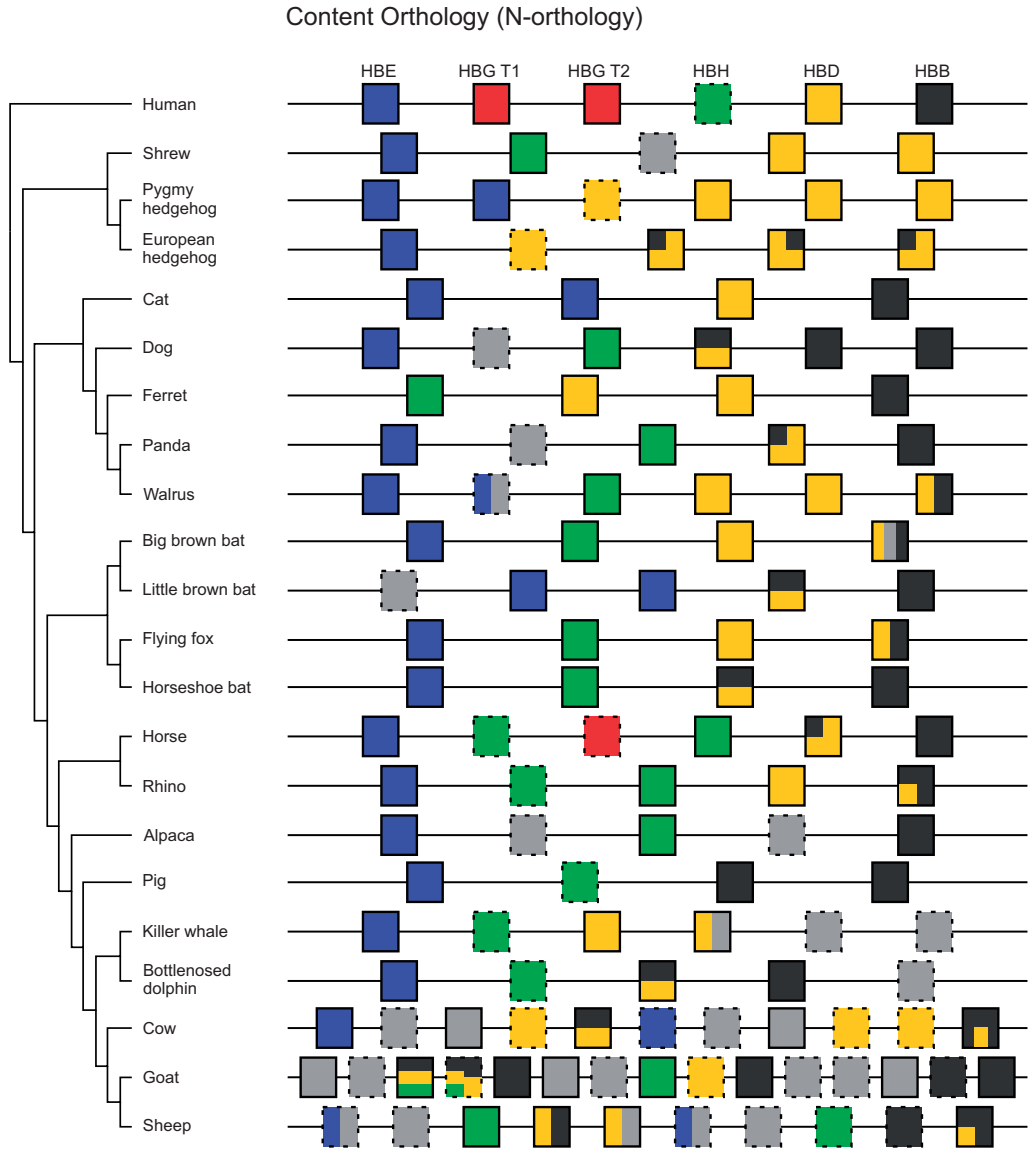

Fig. 7.—Orthologous relationships among β-like globin genes, inferred from pairwise comparisons between the human β-globin cluster, which was used as reference, and a subset of homologous clusters from 21 laurasiatherian species. CHAP2 (Song et al. 2012) was used to assign orthology and to identify cases of interparalog gene conversion. The diagram shows inferred patterns of "content orthology" (N-orthology), which accounts for interparalog gene conversion. Note that gene annotations are not used to make these inferences. Genes that have multiple putative human orthologs are split with a horizontal line, such as the fourth gene in the dog cluster, whereas genes with portions showing discordant affinities are split with vertical lines, such as the last gene in the flying fox cluster. Boxes with dashed lines represent pseudogenes. CHAP2 identified a pseudogene in the sheep cluster (second from the right) that was not identified in our phylogenetic survey. Light gray shading denotes unassigned orthology. The β-like globin genes are shown in the correct order in each cluster (5′–3′), but the gene lengths and intergenic distances are not drawn to scale.

been retained, the newly created *HBB*/*HBD* fusion gene eventually supplanted the parental *HBD* and *HBB* genes, thereby assuming primary (or exclusive) responsibility for synthesizing the β-type subunits of adult and fetal Hb. Among those taxa that have inherited a chimeric "anti-Lepore" duplication, the Canoidea represents the only taxon that has retained intact copies of the parental *HBD* and *HBB* genes along with the *HBB*/*HBD* fusion gene.

Determinants of Relative Expression Levels of HBB-Like and HBD-Like Genes

To assess whether high-level expression requires *HBB*-like promoter sequence, we identified proximal *cis*-regulatory elements within approximately 200 bp of the initiation codon of each *HBB* and *HBD* gene (Myers et al. 1986; Ebb et al. 1998; Ristaldi et al. 1999). We then determined whether

the products of these genes are incorporated into functional Hb tetramers by matching conceptual translations of the coding sequences to the primary structures of β-type Hb subunits that were independently derived via peptide or mRNA sequencing (supplementary table S3, Supplementary Material online). With one notable exception (walrus), genes that possess intact CACCC and CCAAT elements appear to be primarily responsible for encoding the β-type Hb subunits. Losses of these motifs are associated with the downregulation of *HBB* (e.g., in felids and galago), whereas secondary reacquisitions of these motifs are associated with the upregulation of *HBD* (e.g., in rhinoceros; fig. 5). These findings demonstrate the importance of these *HBB*-like regulatory elements for gene expression.

Characterization of proximal cis-regulatory elements of the laurasiatherian *HBD*- and *HBB*-like genes revealed several additional cases of *HBB* → *HBD* gene conversion where the donor sequence included elements of the 5′-upstream regulatory region (fig. 5) but were too short to be identified in our phylogenetic analysis. These conversion events (which included ~130 bp of upstream sequence in each case) extended far enough to restore the CCAAT promoter element, thereby promoting the upregulation of *HBD*. Consequently, in a surprisingly large number of laurasiatherian taxa, *HBD* genes with *HBB*-like promoter elements encode the β-type chains of 20–100% of adult Hb (fig. 5).

Discussion

Results of our comparative genomic and phylogenetic analysis of the laurasiatherian β-globin gene cluster led to the discovery of independently derived *HBB*/*HBD* fusion genes arising from unequal crossing-over in three distinct lineages: Eulipotyphlans, carnivores, and cetaceans (table 1 and figs. 2 and 4). Additionally, a functionally intact *HBB*/*HBD* fusion gene with *HBB*-like proximal cis-regulatory elements originated via gene conversion in the microchiropteran bat genus *Myotis*. Numerous similar, but shorter upstream conversion events were also apparent in the ancestors of shrews, bats, carnivores, and rhinoceros. The availability of independently derived primary structures of β-chain Hbs from representatives of each taxon confirmed that the products of the resulting chimeric *HBB*/*HBD* fusion genes are incorporated into fully functional Hb tetramers at markedly higher levels than human *HBD* (fig. 5). Our analysis also revealed a particularly interesting case of concerted evolution between the *HBD* and *HBB* genes of felids. During postnatal life, the *HBD* and *HBB* genes encode 60–70% and 30–40% of total Hb, respectively (Abbasi and Braunitzer 1985). Both β-type Hb isoforms are unusual in that their oxygen affinities are not allosterically regulated by the intraerythrocytic effector 2,3-diphosphoglycerate (DPG). This insensitivity to DPG is due to a His → Phe substitution at position 2 in the β-type globin chain (Perutz and Imai 1980), a substitution shared by both *HBD* and *HBB* due to interparalog gene conversion. Consequently, Hb isoforms that incorporate β-chain products of either *HBD* or *HBB* have similar modes of allosteric regulation.

In addition to these examples involving mammalian β-like Hb genes, recent comparative genomic studies have revealed that chimeric gene fusions and domain-shuffling events have contributed to the evolution of novel protein functions in a number of more ancient members of the globin gene superfamily in metazoans (Hoffmann et al. 2012; Hoogewijs et al. 2012). Given the increasingly well-documented role of chimeric fusion genes in the evolution of novel protein functions (Patthy 2003), it is important to understand the genetic and evolutionary mechanisms that contribute to their initial fixation and subsequent retention in the genome.

Dispensability of the HBD Gene Varies Among Lineages

The three β-like globin genes that exhibit the highest rates of turnover, and which are most frequently involved in interparalog gene conversion—the embryonic *HBG* and *HBH* genes and the late-expressed *HBD* gene—are located in the center of the gene cluster. The chromosomal interval between the *HBE* gene at the 5′-end of the cluster and the *HBB* gene at the 3′-end can be viewed as a "genomic revolving door" (Demuth et al. 2006) of gene gain, gene loss, and gene fusion. During the evolution of placental mammals, the *HBD* gene has undergone an especially high rate of gene deletion and inactivation, and it has been repeatedly converted by the *HBB* gene (especially at its 5′-end) in rodents, lagomorphs, and primates (Jeffreys et al. 1982; Martin et al. 1983; Hardies et al. 1984; Hardison and Margot 1984; Hoffmann et al. 2008b; Opazo et al. 2008a, 2008b, 2009) in addition to many laurasiatherian taxa included in this study. *HBD* is not expressed in Old World monkeys (Martin et al. 1980), but in hominoids and New World monkeys that have retained a transcriptionally active copy of *HBD*, $\alpha_2\delta_2$ isoforms account for only 1–6% of total Hb in definitive erythrocytes (Boyer et al. 1971; Spritz and Giebel 1988). Such patterns have fostered the impression that *HBD* represents a vestigial gene that has been occasionally resurrected by *HBB* → *HBD* gene conversion that partially restored promoter function (Tagle et al. 1991; Martin et al. 1983; Hardies et al. 1984). As stated by Hardies et al. (1984, p. 3755): "The overall poor evolutionary performance of δ-like genes among mammals suggests that the proto-δ was already destined for disposal prior to the mammalian radiation." However, this view regarding the dispensability of *HBD* was primarily based on data from members of one particular mammalian clade, Euarchontoglires, which includes disproportionately well-studied taxa such as primates and rodents.

Retention of HBD Genes and Pseudogenes

Duplicated genes can be selectively retained in the genome either because evolved functional differences and/or expression differences between the two paralogs are advantageous

or because the loss of subfunctionalized paralogs is deleterious (Force et al. 1999; Zhang 2003; Hahn 2009; Innan and Kondrashov 2010). In humans and other simian primates, there is no evidence to suggest any functionally significant division of labor between the major $\alpha_2\beta_2$ Hb isoform (HbA) and the minor $\alpha_2\delta_2$ Hb isoform (HbA_2) with respect to blood-oxygen transport (Steinberg and Adams 1991; Schechter 2008). In humans, HbA_2 accounts for less than 3% of total adult Hb (Boyer et al. 1971), so any differences in oxygenation properties would have negligible consequences. Any benefit of retaining an intact copy of *HBD* (even if transcriptionally inactive) may relate to incidental position effects on the transcriptional regulation of other prenatally and postnatally expressed β-like globin genes (Moleirinho et al. 2013). Consistent with this hypothesis, results of recent chromosome conformational analyses suggest that *HBD* and the adjacent *HBB* pseudogene may have a regulatory role in maintaining a chromatin conformational state that permits long-range interactions with the downstream locus control region (Sanyal et al. 2012).

Evolutionary Fates of Chimeric Fusion Genes Are Influenced by Their Recombinational Origins

In addition to documenting that the adult Hbs of several laurasiatherian taxa incorporate β-chain products of chimeric *HBB/HBD* fusion genes, our results also indicate that the retention and ascendancy of genes with chimeric coding sequence requires the retention of *HBB*-like promoter sequence arising from unequal cross-over events or the secondary acquisition of *HBB*-like promoter sequence via gene conversion. Available data suggest that each of the expressed *HBB/HBD* genes has *HBB*-like upstream sequence and, consequently, *HBB*-like proximal cis-regulatory elements (fig. 5). Thus, previous authors such as Hardies et al. (1984) appear to have been correct about the importance of retaining *HBB*-like promoters. The well-documented difference in the efficacy of *HBB* and *HBD* promoters (Poncz et al. 1983; Antoniou and Grosveld 1990) provides a logical explanation for why a disproportionate number of chimeric *HBB/HBD* fusion genes have been retained relative to the reciprocal *HBD/HBB* fusion gene.

It is clear that genes with *HBD*-like intron 2 sequence are only expressed if they have *HBB*-like upstream flanking sequence. Primates, lagomorphs, and rodents, the only groups that express $\alpha_2\delta_2$ Hb isoforms in appreciable quantities have *HBD* genes that have been partially overwritten by *HBB*-derived gene conversion at the 5′-end, such that the conversion tract spans the upstream activating sequence and basal promoter (Donze et al. 1996; Tang et al. 1997; Hardison 2001). Gene conversion from *HBB* potentially restores the CCAAC box to CCAAT and donates the proximal CACCC box, which is a key-binding site for Zn-finger transcription factors including the erythroid Krüppel-like factor (Miller and Bieker 1993; Hardison 2001). Our results suggest that similar upstream conversion events were also responsible for upregulating *HBD* expression in numerous laurasiatherian lineages. These findings contribute to a growing awareness of the importance of interparalog gene conversion as a mechanism for generating variation in gene function (Chen et al. 2007; Casola et al. 2012).

In addition to differences in expression levels of the reciprocal *HBB/HBD* and *HBD/HBB* fusion genes, data from human deletion and duplication mutants indicate that the concomitant changes in gene copy number can perturb dosage balance and can alter the Hb isoform composition in circulating erythrocytes (Saller et al. 2012). Given the evidence that the coding sequence of *HBD* may have been under less stringent functional constraints than *HBB* during most of mammalian evolution (Hardies et al. 1984), and given the evidence for distinct structural and functional properties of Hbs with δ-like chains (Vasudevan and McDonald 1998), duplications that increase the proportion of $\alpha_2(\beta/\delta)_2$ Hb isoforms may be expected to have functional consequences for blood-oxygen transport and other aspects of erythrocyte function in laurasiatherian mammals. An obvious prediction is that any unusual properties of $\alpha_2(\beta/\delta)_2$ Hbs will be attributable to amino acid substitutions in the C-terminal, *HBD*-encoded segment of the β-type subunit that occurred during a prior history of relaxed functional constraint.

Supplementary Material

Supplementary tables S1–S3 and figure S1 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank three anonymous reviewers for helpful comments and they thank Giltae Song and Webb Miller for assistance in running and interpreting the results from CHAP2. This work was supported by grants from the National Institutes of Health (R01 HL087216 and HL087216-S1 to J.F.S.), the National Science Foundation (IOS-0949931 to J.F.S. and EPS-0903787 to F.G.H.), and Natural Sciences and Engineering Research Council of Canada Discovery and Accelerator Supplement awards to K.L.C.

Literature Cited

Abbasi A, Braunitzer G. 1985. The primary structure of hemoglobins from the domestic cat (*Felis catus*, Felidae). Biol Chem Hoppe Seyler. 366: 699–704.

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

Antoniou M, Grosveld F. 1990. Beta-globin dominant control region interacts differently with distal and proximal promoter elements. Genes Dev. 4:1007–1013.

Boyer SH, et al. 1971. Primate hemoglobins: some sequences and some proposals concerning the character of evolution and mutation. Biochem Genet. 5:405–448.
Burge C, Karlin S. 1997. Prediction of complete gene structures in human genomic DNA. J Mol Biol. 268:78–94.
Campbell KL, Signore AV, Harada M, Weber RE. 2012. Molecular and physicochemical characterization of hemoglobin from the high-altitude Taiwanese brown-toothed shrew (*Episoriculus fumidus*). J Comp Physiol B. 182:821–829.
Campbell KL, et al. 2010. Molecular basis of a novel adaptation to hypoxic hypercapnia in a strictly fossorial mole. BMC Evol Biol. 10:214.
Casola C, Zekonyte U, Phillips AD, Cooper DN, Hahn MW. 2012. Interlocus gene conversion events introduce deleterious mutations into at least 1% of human genes associated with inherited disease. Genome Res. 22:429–35.
Chen JM, Cooper DN, Chuzhanova N, Férec C, Patrinos GP. 2007. Gene conversion: mechanisms, evolution and human disease. Nat Rev Genet. 8:762–775.
Corduso-Moreira MM, Long M. 2012. The origin and evolution of new genes. In: Anisimova M, editor. Evolutionary genomics: statistical and computational methods. Methods in molecular biology. Vol. 856. New York: Springer. p. 161–186.
Demuth JP, De Bie T, Stajich JE, Cristianini N, Hahn MW. 2006. The evolution of mammalian gene families. PLoS One 1:e85.
Donze D, Jeancake PH, Townes TM. 1996. Activation of *delta-globin* gene expression by erythroid Kruppel-like factor: a potential approach for gene therapy of sickle cell disease. Blood 88:4051–4057.
Ebb D, et al. 1998. Identification of upstream regulatory elements that repress expression of adult beta-like globin genes in a primitive erythroid environment. Blood Cells Mol Dis. 24:356–369.
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.
Force A, et al. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545.
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.
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.
Hahn MW. 2009. Distinguishing among evolutionary models for the maintenance of gene duplicates. J Hered. 100:605–617.
Hardies SC, Edgell MH, Hutchison CA. 1984. Evolution of the mammalian β-globin gene cluster. J Biol Chem. 259:3748–3756.
Hardison RC. 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 RC. 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. p. 95–115.
Hardison RC. 2012. Evolution of hemoglobin and its genes. Cold Spring Harb Perspect Med. 2:a011627.
Hardison RC, 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 RC, Margot JB. 1984. Rabbit globin pseudogene *ψβ2* is a hybrid of *δ*- and *β*-globin sequences. Mol Biol Evol. 1:302–316.
Hoffmann FG, Opazo JC, Storz JF. 2008a. Rapid rates of lineage-specific gene duplication and deletion in the α-globin gene family. Mol Biol Evol. 25:591–602.
Hoffmann FG, Opazo JC, Storz JF. 2008b. New genes originated via multiple recombinational pathways in the β-globin gene family of rodents. Mol Biol Evol. 25:2589–2600.
Hoffmann FG, Storz JF. 2007. The α^{D}-globin gene originated via duplication of an embryonic α-like globin gene in the ancestor of tetrapod vertebrates. Mol Biol Evol. 24:1982–1990.
Hoffmann FG, Storz JF, Gorr TA, Opazo JC. 2010. Lineage-specific patterns of functional diversification in the α- and β-globin gene families of tetrapod vertebrates. Mol Biol Evol. 27:1126–1138.
Hoffmann FG, et al. 2012. Evolution of the globin gene family in deuterostomes: lineage-specific patterns of diversification and attrition. Mol Biol Evol. 29:1735–1745.
Holloway K, Lawson VE, Jeffreys AJ. 2006. Allelic recombination and de novo deletions in sperm in the human β-globin gene region. Hum Mol Genet. 15:1099–1111.
Hoogewijs D, et al. 2012. Androglobin: a chimeric globin in metazoans that is preferentially expressed in mammalian testes. Mol Biol Evol. 29: 1105–1114.
Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet. 11:97–108.
Jeffreys AJ, et al. 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.
Jones CD, Begun DJ. 2005. Parallel evolution of chimeric fusion genes. Proc Natl Acad Sci U S A. 102:11373–11378.
Jones CD, Custer AW, Begun DJ. 2005. Origin and evolution of a chimeric fusion gene in *Drosophila subobscura*, *D. madeirensis*, and *D. guanche*. Genetics 170:207–219.
Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. Genome Res. 20:1313–1326.
Katju V. 2012. In with the old, in with the new: the promiscuity of the duplication process engenders diverse pathways for novel gene creation. Int J Evol Biol. 2012:341932.
Katju V. 2013. To the beat of a different drum: determinants implicated in the asymmetric sequence divergence of *Caenorhabditis elegans* paralogs. BMC Evol Biol. 13:73.
Katju V, Lynch M. 2003. The structure and early evolution of recently arisen gene duplicates in the *Caenorhabditis elegans* genome. Genetics 165:1793–1803.
Katju V, Lynch M. 2006. On the formation of novel genes by duplication in the *Caenorhabditis elegans* genome. Mol Biol Evol. 23:1056–1067.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30:772–780.
Kellis M, et al. 2014. Defining functional DNA elements in the human genome. Proc Natl Acad Sci U S A. 111:6131–6138.
Koop BF, et al. 1989. Tarsius *δ*- and *β*-globin genes: conversions, evolution, and systematics. J Biol Chem. 264:68–79.
Long ME, Betran E, Thornton K, Wang W. 2003. The origin of new genes: glimpses from the young and old. Nat Rev Genet. 4:865–875.
Martin SL, Vincent KA, Wilson AC. 1983. Rise and fall of the *δ*-globin gene. J Mol Biol. 164:513–528.
Martin SL, Zimmer EA, Kan YW, Wilson AC. 1980. Silent *δ*-globin gene in Old World monkeys. Proc Natl Acad Sci U S A. 77:3563–3566.

Meredith RW, et al. 2011. Impacts of the cretaceous terrestrial revolution and KPg extinction on mammal diversification. Science 334: 521–524.
Miller IJ, Bieker JJ. 1993. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Krüppel family of nuclear proteins. Mol Cell Biol. 13: 2776–2786.
Moleirinho A, et al. 2013. Evolutionary constraints in the β-globin gene cluster: the signature of purifying selection at the δ-globin (*HBD*) locus and its role in developmental gene regulation. Genome Biol Evol. 5: 559–571.
Myers RM, Tilly K, Maniatis T. 1986. Fine structure genetic analysis of a beta-globin promoter. Science 232:613–618.
Nery MF, González DJ, Hoffmann FG, Opazo JC. 2012. Resolution of the laurasiatherian phylogeny: evidence from genomic data. Mol Phylogenet Evol. 64:685–689.
Olivieri NF, Weatherall DJ. 2001. Clinical aspects of beta thalassemia. In: Steinberg MH, Forget BG, Higgs DR, Nagel RL, editors. Disorders of hemoglobin: genetics, pathophysiology, and clinical management. Cambridge: Cambridge University Press. p. 277–341.
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 U S A. 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 U S A. 105:12950–12955.
Opazo JC, Sloan A, Campbell KL, Storz JF. 2009. Origin and ascendancy of a chimeric fusion gene: the *β*/*δ*-globin gene of paenungulate mammals. Mol Biol Evol. 26:1469–1478.
Patthy L. 2003. Modular assembly of genes and the evolution of new functions. Genetica 118:217–231.
Perutz MF, Imai K. 1980. Regulation of oxygen affinity of mammalian haemoglobins. J Mol Biol. 136:183–191.
Poncz M, Schwartz E, Ballantine M, Surrey S. 1983. Nucleotide sequence analysis of the delta beta-globin gene region in humans. J Biol Chem. 258:11599–11609.
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.
Rachmilewitz E, Schrier S. 2001. Pathophysiology 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. 233–251.
Ristaldi MS, et al. 1999. Activation of the *δ*-globin gene by the *β*-globin gene CACCC motif. Blood Cells Mol Dis. 25:193–209.
Rogers RL, Bedford T, Hartl DL. 2009. Formation and longevity of chimeric and duplicate genes in *Drosophila melanogaster*. Genetics 181: 313–322.
Rogers RL, Bedford T, Hartl DL. 2010. Adaptive impacts of the chimeric gene *Quetzlcoatl* in *Drosophila melanogaster*. Proc Natl Acad Sci U S A. 107:10943–10948.
Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.
Runck AM, Moriyama H, Storz JF. 2009. Evolution of duplicated *β*-globin genes and the structural basis of hemoglobin isoform differentiation in *Mus*. Mol Biol Evol. 26:2521–2532.
Runck AM, Weber RE, Fago A, Storz JF. 2010. Evolutionary and functional properties of a two-locus *β*-globin polymorphism in Indian house mice. Genetics 184:1121–1131.
Saller E, et al. 2012. Comparison of two known chromosomal rearrangements in the δβ-globin complex with identical DNA breakpoints but causing different HbA_2 levels. Hemoglobin 36: 177–182.
Sanyal A, Lajoie BR, Jain G, Dekker J. 2012. The long-range interaction landscape of gene promoters. Nature 489:109–113.
Schechter AN. 2008. Hemoglobin research and the origins of molecular medicine. Blood 112:3927–3938.
Schimenti JC, Duncan CH. 1985. Structure and organization of the bovine *β*-globin genes. Mol Biol Evol. 2:514–525.
Schwartz S, et al. 2000. PipMaker—a web server for aligning two genomic DNA sequences. Genome Res. 10:577–586.
Schwartz S, et al. 2003. MultiPipMaker and supporting tools: alignments and analysis of multiple genomic DNA sequences. Nucleic Acids Res. 31:3518–3524.
Signore AV, Stetefeld J, Weber RE, Campbell KL. 2012. Origin and mechanism of thermal insensitivity in mole hemoglobins: a test of the "additional" chloride binding site hypothesis. J Exp Biol. 215:518–525.
Song G, et al. 2012. Revealing mammalian evolutionary relationships by comparative analysis of gene clusters. Genome Biol Evol. 4:586–601.
Spritz RA, Giebel LB. 1988. The structure and evolution of the spider monkey δ-globin gene. Mol Biol Evol. 5:21–29.
Steinberg M, Adams JG. 1991. Hemoglobin A_2: origin, evolution, and aftermath. Blood 78:2165–2177.
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, Natarajan C, Cheviron ZA, Hoffmann FG, Kelly JK. 2012. Altitudinal variation at duplicated *β*-globin genes in deer mice: effects of selection, recombination, and gene conversion. Genetics 190: 203–216.
Storz JF, Opazo JC, Hoffmann FG. 2011. Phylogenetic diversification of the globin gene superfamily in chordates. IUBMB Life 63:313–322.
Storz JF, Opazo JC, Hoffmann FG. 2013. Gene duplication, genome duplication, and the functional diversification of vertebrate globins. Mol Phylogenet Evol. 66:469–478.
Storz JF, Runck AM, Moriyama H, Weber RE, Fago A. 2010. Genetic differences in hemoglobin function between highland and lowland deer mice. J Exp Biol. 213:2565–2574.
Storz JF, et al. 2007. Complex signatures of selection and gene conversion in the duplicated globin genes of house mice. Genetics 177:481–500.
Storz JF, et al. 2009. Evolutionary and functional insights into the mechanism underlying high-altitude adaptation of deer mouse hemoglobin. Proc Natl Acad Sci U S A. 106:14450–14455.
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.
Tang DC, Ebb D, Hardison RC, Rodgers GP. 1997. Restoration of the CCA AT box or insertion of the CACCC motif activates [corrected] *delta*-globin gene expression. Blood 90:421–427.
Tatusova TA, Madden TL. 1999. BLAST 2 sequences, a new tool for comparing protein and nucleotide sequences. FEMS Microbiol Lett. 174: 247–250.
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 U S A. 81:6589–6593.
Vasudevan G, McDonald MJ. 1998. Analysis of the global architecture of Hemoglobin A_2 by heme binding studies and molecular modeling. J Protein Chem. 17:319–327.
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. 2003. Evolution by gene duplication: an update. Trends Ecol Evol. 18:292–298.

Associate editor: Ross Hardison