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

Michael J. Gaudry1, Jay F. Storz2, Gary Tyler Butts3, Kevin L. Campbell1, and Federico G. Hoffmann3,4,*

1Department of Biological Sciences, University of Manitoba, Winnipeg, MB, Canada
2School of Biological Sciences, University of Nebraska, Lincoln
3Department of Biochemistry, Molecular Biology, Entomology, and Plant Pathology, Mississippi State University
4Institute for Genomics, Biocomputing and Biodiversity, 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 postnataally 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/\textit{HBB} fusion genes due to thalassemia pathologies associated with the \textit{HBD}/\textit{HBB} “Lepore” deletion mutant in humans. Here, we report a comparative genomic analysis of the mammalian β-globin gene cluster, which revealed that chimeric \textit{HBB}/\textit{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 \textit{HBD} and/or \textit{HBB} genes have typically been inactivated or deleted, so that the newly created \textit{HBB}/\textit{HBD} fusion gene is primarily responsible for synthesizing the β-type subunits of adult and fetal hemoglobin (Hb). Contrary to conventional wisdom that the \textit{HBD} gene is a vestigial relict that is typically inactivated or expressed at negligible levels, we show that \textit{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 \textit{HBD}-like coding sequence requires the secondary acquisition of \textit{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. 1A). 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
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 (HBBH), 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; Pritchito 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

![Figure 1](https://example.com/fig1.png)

**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 \( \beta \)-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. 1A).

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 \( \beta \)-chain products of the chimeric fusion gene. The lower abundance of the Hb Lepore isoform is mainly due to a dosage imbalance results in insoluble aggregations of oxidized \( \alpha \)-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).

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 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.

Materials and Methods

Annotation of Genomic Sequences

We obtained genomic sequences containing the \( \beta \)-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 \( \beta \)-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.
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; Pycrito 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 pri-mates), 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 HBB; 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 HBB. 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 HBB 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 HBB pseudogene, most laurasiatherians possess functional copies of HBB 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 HBB 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 HBB 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. 4A). 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. 4B–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. 4C).

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 HBD or HBB 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.
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 regions
Table 1
Patterns of Sequence Chimerism in a Sample of 70 Late-Expressed β-Like Globin Genes in Laurasiatherian Mammals

<table>
<thead>
<tr>
<th>Cross-Over Type</th>
<th>Chimeric Pattern (5′-Intron 2-3)</th>
<th>Genes</th>
<th>Pseudogenes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anti-Lepore</td>
<td>β-β-β</td>
<td>30</td>
<td>—</td>
</tr>
<tr>
<td>Lepore</td>
<td>β-β-δ</td>
<td>1</td>
<td>Cetacea</td>
</tr>
<tr>
<td></td>
<td>(Turisiops)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>β-δ-δ</td>
<td>16</td>
<td>Carnivora,</td>
</tr>
<tr>
<td></td>
<td>Chiroptera</td>
<td></td>
<td>Eulipotyphla</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lepore</td>
<td>δ-β-β</td>
<td>1</td>
<td>Cetartiodactyla</td>
</tr>
<tr>
<td></td>
<td>(Cow)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>δ-β-δ</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Cetartiodactyla</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>δ-δ-δ</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>

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 HBB and HBD genes (see text for details). The reciprocal HBB/HBD ("1H-β-β" and "1H-β-δ") and HBB-HBD ("1H-β-δ" and "1H-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 "1H-β-δ" and "1H-δ-δ" genes represent cases where each of the three noncoding segments match the corresponding segments of the human HBB and HBD genes, respectively.

Independent Origins of Chimeric HBB/HBD Fusion Genes

The HBB/HBD fusion genes of the bottlenose dolphin (Turisiops 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. 4C, 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 boids. 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 HBB 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 vampyrus), and big brown bat (Eptesicus fuscus).
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).
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 eulipotyphlans, 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

![Diagram](http://gbe.oxfordjournals.org/)

**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.
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 \( \beta \)-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-type 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, ε2ε2 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 the ε-like genes among mammals suggests that the proto-ε 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...
Repeated Evolution of Chimeric β-Globins in Laurasiatheria

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 αβ 2 Hb isofrom (HbA) and the minor αβ 2 Hb isofrom (HbA2) with respect to blood-oxygen transport (Steinberg and Adams 1991; Schechter 2008). In humans, HbA2 accounts for less than 3% of total adult Hb (Boyer et al. 1971), so any differences in oxygenation properties would have negligible consequences. Any beneficial in 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 ascendency 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 Grosfeld 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 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 αδ(βδ) Hbs 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 αδ(βδ) Hbs will be attributable to amino acid substitutions in the C-terminal, HBD-encoded segment of the δ-chain 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 Gitaie 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


