Rapid Rates of Lineage-Specific Gene Duplication and Deletion in the \(\alpha\)-Globin Gene Family

Federico G. Hoffmann, Juan C. Opazo,\(^1\) and Jay F. Storz

School of Biological Sciences, University of Nebraska

Phylogeny reconstructions of the globin gene families have revealed that paralogous genes within species are often more similar to one another than they are to their orthologous counterparts in closely related species. This pattern has been previously attributed to mechanisms of concerted evolution such as interparalog gene conversion that homogenize sequence variation between tandemly duplicated genes and therefore create the appearance of recent common ancestry. Here we report a comparative genomic analysis of the \(\alpha\)-globin gene family in mammals that reveal a surprisingly high rate of lineage-specific gene duplication and deletion via unequal crossing-over. Results of our analysis reveal that patterns of sequence similarity between paralogous \(\alpha\)-like globin genes from the same species are only partly explained by concerted evolution between preexisting gene duplicates. In a number of cases, sequence similarity between paralogous sequences from the same species is attributable to recent ancestry between the products of de novo gene duplications. As a result of this surprisingly rapid rate of gene gain and loss, many mammals possess \(\alpha\)-like globin genes that have no orthologous counterparts in closely related species. The resultant variation in gene copy number among species may represent an important source of regulatory variation that affects physiologically important aspects of blood oxygen transport and aerobic energy metabolism.

Introduction

Phylogeny reconstructions of gene family evolution often reveal that paralogous genes within species are more similar to one another than they are to their orthologous counterparts in closely related species. This pattern is a hallmark of concerted evolution and is typically attributed to the homogenizing effects of interparalog gene conversion or unequal crossing-over (Zimmer et al. 1980; Ohta 1984, 1990, 2000). Gene conversion involves a nonreciprocal recombination event between paralogous sequences and is thought to be the most important mechanism of concerted evolution in small multigene families (Dover 1982; Nagylaki and Peters 1982; Nagylaki 1984a, 1984b; Ohta 1990). Unequal crossing-over is a reciprocal recombination event that produces a sequence duplication on 1 chromatid or chromosome and a corresponding deletion in the other. Repeated rounds of unequal crossing-over can result in concerted evolution in cases where 1 paralogous sequence is propagated at the expense of other tandemly duplicated loci, thereby progressively homogenizing sequence variation among members of the gene family (Ohta 1980, 1984; Gojobori and Nei 1984; Li et al. 1985). The role of both gene conversion and unequal crossing-over in homogenizing sequence variation among tandemly duplicated genes has been especially well documented in the globin gene families (Jeffreys 1979; Slightom et al. 1982; Czelusniak et al. 1982), and in amniote vertebrates the 2 gene families are located on different chromosomes. Most marsupial and placental mammals possess 4 different \(\alpha\)-like globin genes: \(\xi\)-globin (HBZ), \(\eta\)-globin (HBK), \(\alpha\)-globin (HBA), and \(\delta\)-globin (HBQ). The HBZ and HBA genes both encode \(\alpha\)-chain subunits of hemoglobin, but they are expressed at different stages of development. HBZ is expressed in primitive erythroid cells in the yolk sac during the earliest stages of embryogenesis, and HBA is expressed in definitive erythrocytes during fetal development and postnatal life (Higgs et al. 1989; Hardison 2001; Nagel and Steinberg 2001). In contrast to the HBZ and HBA genes, the HBK and HBQ genes do not appear to encode subunit polypeptides of hemoglobin in mammals, and their functions have yet to be illuminated. The duplication events that produced the HBZ, HBK, and HBA genes predated the origin of tetrapod vertebrates (Goodman et al. 1975, 1987; Hoffmann and Storz 2007), whereas the HBQ gene appears to be the product of a mammal-specific duplication of the HBA gene (Cooper et al. 2005).

The majority of mammals studied to date possess either 1 or 2 functional copies of HBZ and either 2 or 3 functional copies of HBA. It has often been assumed that the same tandemly duplicated HBZ and HBA genes were inherited from the common ancestor of all mammals (Zimmer et al. 1980; Flint et al. 1988; Hardison 2001). According to this scenario, the 5' HBA gene in 1 species is assumed to be orthologous to the 5' HBA gene of all other species and likewise for the other HBA and HBZ paralogs (Flint et al. 1988; Hardison 2001). The fact that paralogous \(\alpha\)-like globin genes within the genome of the same species are often identical or nearly identical in sequence has typically been attributed to concerted evolution (Zimmer et al. 1980; Liebhaber et al. 1981; Proudfoot et al. 1982; Michelson and Orkin 1983). According to this explanation, the homogenization of sequence variation between paralogous \(\alpha\)-globin genes erases phylogenetic history and creates the appearance of recent common ancestry (Zimmer et al. 1980; Liebhaber et al. 1981; Proudfoot et al. 1982; Michelson and Orkin 1983; Higgs et al. 1989; Hardison 2001). As stated by Graur and Li (2000, p. 314) in explaining the observed sequence similarity between paralogous
HBA genes in humans and other mammals: “...one had to assume either that multiple gene duplication events occurred independently in many evolutionary lineages or that the two genes are quite ancient, having been duplicated once in the common ancestor of these organisms, but their antiquity was subsequently obscured by concerted evolution. Ultimately, the most parsimonious solution was to choose the latter alternative.” Although this interpretation of the observed phylogenetic patterns may be the most parsimonious, results of our comparative genomic analysis reveal that it is not completely correct. Here we report a detailed analysis of sequence variation in coding regions and flanking regions of mammalian α-like globin genes that reveals a surprisingly high rate of lineage-specific gene duplication and deletion via unequal crossing-over. Results of our analyses reveal that observed patterns of sequence similarity between paralogous HBZ and HBA genes are only partly explained by concerted evolution. In many cases, the appearance of recent common ancestry between paralogous sequences is real as new α-like globin genes have originated multiple times independently in different lineages of placental mammals.

The objective of this study was to assess the relative importance of concerted evolution and birth-and-death evolution in shaping the genomic structure of the α-globin gene family in mammals. Specifically, we used genomic sequence data 1) to characterize the genomic structure of the mammalian α-globin gene family, 2) to assign orthologous and paralogous relationships among duplicate copies of α-like globin genes, and 3) to assess whether sequence similarity between paralogs within the same species’ genome is typically attributable to concerted evolution between preexisting gene duplicates or recent ancestry between duplicated genes that originated independently in different lineages.

Materials and Methods
DNA Sequence Data and Bioinformatic Analyses

Genomic sequences that spanned all or most of the α-globin gene cluster were identified in either GenBank or Ensembl databases by BlastN alignment to known α-like globin sequences. When possible, we focused on sequences from a single genomic contig, genomic scaffold, or full chromosome, depending on the nature of the available data. The basic annotation was derived from the database records in most cases, but we also identified globin genes in unannotated sequences using GENSCAN (Burge and Karlin 1997) and by comparing known exon sequences with genomic contigs using the program Blast2 sequences version 2.2 (Tatusova and Madden 1999) from the National Center for Biotechnology Information Blast suite (http://www.ncbi.nlm.nih.gov/blast). Annotated genes were considered to be functional when they met the following criteria: there were no premature stop codons, there were no frameshift mutations, and a stop codon was present at codon position 42 of the third exon. Because of incomplete sequence coverage of the gene cluster, there were some genomic sequences in our data set for which we could not ascertain the full extent of conserved synteny. These include genomic sequences from the cat (Felis domesticus) and the stripe-face dunnart (Sminthopsis macroura). Genomic sequences were masked using RepeatMasker (http://www.repeatmasker.org), and genomic sequence alignments were conducted using Pipmaker (Schwartz et al. 2000), Multipipmaker (Schwartz et al. 2003), and Mulan (Ovcharenko et al. 2005). In order to identify tandemly duplicated genes or sets of genes, we used percent identity plots to identify short chromosomal regions that were locally alignable to 1 or more additional regions within the same genomic contig. For the intragenomic dot plot analyses, we focused on contigs that included 50 kb of flanking sequence upstream and downstream of the α-globin gene cluster.

Phylogeny Reconstruction

We explored phylogenetic relationships of α-globin genes at several levels. In all cases, sequences were aligned using ClustalX (Thompson et al. 1997). We inferred phylogenetic relationships in a maximum likelihood framework using Treefinder version June 2007 (Jobb et al. 2004) and assessed support for the nodes with 1,000 bootstrap pseudoreplicates. In analyses restricted to protein-coding sequences, an independent model of nucleotide substitution was used for each codon position. Phylogenetic results were robust to variation in the model of nucleotide substitution selected; here, we report results obtained under the general time-reversible model (Rodriguez et al. 1990) in which rate variation followed a discrete gamma distribution (GTR + Γ). Due to the fact that intronic sequences from distantly related species were often unalignable, we restricted the analysis to coding sequence. We followed a similar strategy to reconstruct phylogenetic trees for all putatively functional HBZ and HBA genes. Phylogeny reconstructions that deviated from the expected species phylogeny were investigated using the approximately unbiased test (Shimodaira 2002), as implemented in Treefinder.

To reconstruct the history of gene duplications and deletions in the α-globin gene cluster of primates, we compared the coding sequences of the genes and the corresponding upstream and downstream flanking regions. Because gene conversion tracts are often restricted to coding regions (Chen et al. 2007), in many cases orthologous relationships between duplicated genes can still be reliably inferred by examining flanking sequence that lies outside of gene conversion tracts (Hardison and Gelines 1986; Hardison and Miller 1993; Storz, Baze, et al. 2007). In order to identify interparalog conversion tracts in primates that possess 3 or more copies of HBA, we used the program GENECONV (Sawyer 1989) with the G-scale parameter (mismatch penalty) set to 2.0. For the HBZ and HBA genes, we conducted phylogeny reconstructions on 3 different partitions of the alignment: the coding sequence, upstream flanking sequence, and downstream flanking sequence. For the HBZ genes, phylogeny reconstructions were based on a fragment that started 500-bp upstream of the start codon and ended 500-bp downstream of the stop codon. In the case of the HBA genes, the first set of analyses included all primates and was based on an alignment that started 1-kb upstream of the start codon and ended 1-kb downstream of the stop codon. A second analysis focused on resolving
relationships among the HBA genes of anthropoid pri-
mates, the group that includes New World monkeys, Old
World monkeys, and apes. For the anthropoids, we aligned
a fragment that started 2-kb upstream of the start codon and
ended 1-kb downstream of the stop codon.

Results and Discussion
Genomic Sequence Data

We obtained genomic sequences that spanned all or
most of the \( \alpha \)-globin gene cluster of 40 mammalian species
(supplementary table S1, Supplementary Material online).
These genomic contigs ranged in size from 10 to 100 kb.
This sample of genomic sequences included representa-
tives of the 3 subclasses of mammals: Prototheria (monotremes),
Metatheria (marsupials), and Eutheria (placental mam-
mals). The sample of placental mammals included repre-
sentatives of each of the 4 superorders: Afrotheria,
Xenarthra, Laurasiatheria, and Euarchontoglires.

Following the nomenclature of Aguileta et al. (2006),
we refer to the \( \zeta \)-globin gene, the \( \alpha^D \)-globin gene, the \( \alpha^A \)-
globin gene, and the \( \theta \)-globin gene, as HBZ, HBK, HBA,
and HBQ, respectively. Because mammalian \( \alpha \)-globin
genes have undergone multiple rounds of duplication
that have resulted in tandemly repeated sets of paralogous
gene copies (Zimmer et al. 1980; Czelusniak et al. 1982;
Proudfoot et al. 1982; Hardison and Gelas 1986; Cheng
et al. 1987; Goodman et al. 1987; Flint et al. 1988, 2001),
we index each duplicated gene with the symbol \( \sim \) fol-
lowed by a number that corresponds to the linkage order
in the 5’ to 3’ orientation (Aguileta et al. 2006).

Genomic Structure of the Mammalian \( \alpha \)-Globin
Gene Cluster

Results of intragenomic dot plot analyses revealed that
tandemly duplicated gene regions were exclusively restricted
to the \( \alpha \)-globin gene cluster (supplementary fig. S1, Supple-
mentary Material online). Genomic sequence comparisons
among monotremes, marsupials, and placental mammals re-
vealed conserved synteny across the entire \( \alpha \)-globin gene
cluster. In representatives of all 3 subclasses of mammals,
the 5’ end of the \( \alpha \)-globin gene cluster is located downstream
of the ortholog of the human \( C16orf35 \) gene and the 3’
end of the gene cluster is located upstream of the ortholog of
the human \( Luc7L \) gene. The 1 notable exception to this pattern is
the house mouse (\( M. \) musculus). In this species, the 5’
end of the \( \alpha \)-globin gene cluster is located on Chromosome
11 but the 3’ end of the cluster, including pseudogene copies
of HBA and HBQ, has been translocated to Chromosome 17
(Flint et al. 2001; Tufarelli et al. 2001).

In nearly all the genomic sequences in our data set that
had complete coverage of the \( \alpha \)-globin gene cluster, the HBZ
and HBQ genes were located at the 5’ and 3’ ends of the
cluster, respectively (fig. 1). As is generally the case in the globin
gene clusters of vertebrates, the embryonic HBZ genes were
located upstream of the adult HBA genes. The only exceptions involved en bloc duplications in the tenrec (\( Echinops \)
telfairi) and the rabbit (\( Oryctolagus cuniculus \), where

![Genomic structure of the \( \alpha \)-globin cluster in mammals. Phylogenetic relationships among mammalian species are based on a loose consensus of recent studies (Murphy et al. 2001, 2007; Hallstrom et al. 2007; Wildman et al. 2007). Diagonal slashes indicate gaps in genomic coverage. Segments containing such gaps were not drawn to scale. Pseudogene fragments containing less than 2 complete exons were not included. The orientation of the clusters is from 5’ (on the left) to 3’ (on the right).](image-url)
HBZ genes in the 3′ duplication block were located downstream of HBA genes in the 5′ duplication block (fig. 1).

We found that all species possess at least 1 functional copy of HBZ and HBA (fig. 1). By contrast, HBK is missing from the genomes of the glires (Rodentia + Lagomorpha) and Afrotherians, and HBQ is missing from the genomes of the shrew (Sorex araneus), the armadillo (Dasypus novemcinctus), and the platypus (Ornithorhyncus anatinus).

In contrast to the HBZ, HBA, and HBQ genes, the HBK gene was never present in more than 1 copy. In our data set, the number of putatively functional genes ranged from 2 in the armadillo (HBZ and HBA) to 8 in the rabbit (4 copies of HBZ, 1 copy of HBA, and 3 copies of HBQ). The observed variation in gene copy number is primarily attributable to tandem duplications of single genes, although there is also evidence for en bloc duplications involving sets of 2–3 closely linked genes. For example, triplication of an ancestral HBZ–HBQ gene pair is evident in the α-globin gene cluster of the rat (Rattus norvegicus) (Storz, Hoffmann, et al. 2008), and the rabbit has several variant copies of an HBZ–HBZ–HBA–HBQ repeat motif, as first reported by Cheng et al. (1987).

Ancestral State of the Mammalian α-Globin Gene Cluster

Based on a comparative analysis of the α-globin gene cluster of marsupials and placental mammals, Cooper et al. (2006) proposed that the α-globin gene cluster in the common ancestor of therian mammals (marsupials + placentals) contained 7 α-like globin genes, in addition to a single copy of ω-globin (HBW) at the 3′ end of the cluster: 5′-HBZ-T1, HBZ-T2, HBK, HBA-T1, HBA-T2, HBA-T3, and HBQ. The HBW gene is a β-like globin gene that has previously been described only in marsupials (Wheeler et al. 2001, 2004; De Leo et al. 2005). The location of the HBW gene at the 3′ end of the α-globin gene cluster reflects the ancestral linkage arrangement of α- and β-like globin genes in the common ancestor of amniote vertebrates (Wheeler et al. 2001, 2004). The availability of genomic sequence from a monotreme taxon, the platypus, provides an opportunity to evaluate the hypothesized structure of the ancestral α-globin gene cluster at the stem of the mammalian radiation. We identified 6 α-like globin genes in the platypus: 5′-HBZ-T1, HBZ-T2, HBK, HBA-T1, HBA-T2, HBA-T3-3′. We also confirmed the presence of HBW at the 3′ end of the cluster (fig. 1). However,
we found no evidence of an HBQ gene in the \( \alpha \)-globin gene cluster of the platypus. Thus, aside from the absence of HBQ, the platypus \( \alpha \)-globin cluster is similar to the ancestral therian \( \alpha \)-globin gene cluster proposed by Cooper et al. (2006). The existence of duplicated copies of HBZ and HBA in the \( \alpha \)-globin gene cluster of monotremes, marsupials, and placental mammals suggests that duplicate copies of both genes may have been present in the common ancestor of all extant mammals. The absence of an HBQ gene in the \( \alpha \)-globin gene cluster of the platypus suggests that the duplication that gave rise to HBQ occurred after the divergence between monotremes and therian mammals. Conversely, the fact that we found no trace of HBW in any of the eutherian mammals examined indicates that the loss of HBW from the 3' end of the \( \alpha \)-globin gene cluster pre-dates the radiation of extant placental mammals. Accordingly, the principle of parsimony suggests the following gene arrangement in the last common ancestor of all extant mammals: 5' HBZ-T1, HBK, HBA-T1, HBA-T2, HBA-T3, HBW-3'. This inferred ancestral gene arrangement is identical to the consensus gene arrangement in extant mammals, except that in placental mammals HBQ has been added and HBW has been lost.

Evolution of Duplicate Copies of HBZ and HBA

In several cases, phylogeny reconstructions of mammalian \( \alpha \)-like globin genes did not recover the expected set of species relationships but deviations from the expected species relationships were not statistically significant. This discordance between the inferred gene trees and the expected species trees is not surprising given the number of informative sites in the alignment relative to the number of taxa. The trees in figure 2 correspond to results of phylogeny reconstructions for the mammalian HBA and HBZ genes where sequences were constrained to match the systematic relationships among the major mammalian sub-classes: (Monotremes (Marsupials, Placentals)).

The phylogenies obtained show the hallmark of concerted evolution: paralogous copies of HBA and HBZ form monophyletic clades within species, to the exclusion of sequences from other species (fig. 2). The HBZ genes of marsupials represent the 1 notable exception to this general pattern (fig. 2A). Within marsupials, the HBZ-T1 and HBZ-T2 paralogs from the tammar wallaby (Macropus eugenii), opossum (Didelphis virginiana), and short-tailed opossum (Monodelphis domestica) are reciprocally monophyletic to one another. In each case, the HBZ-T1 and HBZ-T2 clades both recover the expected species
phylogeny; *Macropus* (*Didelphis, Monodelphis)*). Dot plot comparisons between HBZ paralogs in monotremes, marsupials, and placental mammals are consistent with the inferences drawn from phylogenetic analyses. In monotremes and placental mammals, there are good sequence matches between the paralogous HBZ genes within the same species. However, dot plot comparisons between the HBZ-T1 and HBZ-T2 paralogs of marsupials revealed the presence of a ~1-kb block of nonhomology in the second intron (supplementary fig. S2, Supplementary Material online).

Evolution of the *α*-Globin Gene Cluster in Primates

To investigate mechanisms of gene family evolution in more detail, we focused our analysis of sequence variation on the *α*-globin gene cluster of primates, the taxon for which we have the most genomic sequence data. In addition to the previously characterized human *α*-globin gene cluster on Chromosome 16 (GenBank accession number NG_000006 [Flint et al. 2001]), we characterized the genomic structure of the *α*-globin gene cluster in an additional 14 primate species (3 prosimians, 4 New World monkeys, 3 Old World monkeys, and 4 apes). The primate species represented in our data set possess either 1 or 2 copies of HBZ, 1 copy of HBK, 1–3 copies of HBA, and 1 copy of HBQ (fig. 3).

Because systematic relationships among the species in our study have been the subject of intensive study (Goodman et al. 2005; Opazo et al. 2006), we have a solid phylogenetic framework for reconstructing gains and losses of *α*-like globin genes over the course of primate evolution (fig. 3). As described below, we found that the vast majority of de novo duplications and deletions of *α*-like globin genes in primates can be attributed to unequal crossing-over events. An unequal crossing-over event between 2 chromosomes that both carry a tandemly duplicated pair of genes will produce 2 daughter chromosomes that carry either 1 or 3 copies of the gene, the historical record of this event is written in the pattern of sequence variation in upstream and downstream flanking regions (fig. 4).

Evolution of the Primate *α*-Globin Gene Cluster

The objective of our initial analyses was to assign orthologous relationships among the multiple HBZ and HBA genes found in primates. Although phylogenetic reconstructions based on coding sequence indicate that orthologous relationships among HBZ and HBA genes have been obscured by a history of concerted evolution, the true history of gene duplication and species divergence is revealed by sequence variation in flanking regions (figs. 5 and 6). In the case of the 5′ HBZ genes of primates, gene conversion tracts appear to be restricted to the exons and introns of the genes as phylogenetic analyses based on upstream and downstream flanking sequence recover the expected species relationships with strong bootstrap support. In all primate species that possess multiple HBZ copies, our phylogeny reconstructions reveal evidence for 1:1 orthology among the full set of 5′ HBZ genes and among all the 3′ HBZ genes with the exception of the guereza (*Colobus guereza*; figs. 5 and 7A).

Assigning orthology among the HBA genes was complicated because gene conversion tracts often extend upstream and downstream of the coding region (supplementary table S2, Supplementary Material online). Analyses based on 1 kb of flanking sequence upstream of the start codon strongly suggest that the 5′ HBA gene of prosimians is orthologous to the 5′ HBA pseudogene of most Old World monkeys and apes. Likewise, analyses based on 1 kb of flanking sequence downstream of the stop codon indicate that the 3′ HBA genes of most primates are 1:1 orthologs (figs. 6 and 7B). Interestingly, none of the New World monkeys appear to possess an ortholog of the 5′ HBA gene of prosimians, Old World monkeys, and apes. Based on these data, there are 2 equally parsimonious reconstructions of the ancestral *α*-globin gene cluster of primates. One possibility is that the ancestral *α*-globin gene cluster contained duplicate copies of both HBZ and HBA, and the other possibility that it contained duplicate copies of HBB and triplicate copies of HBA. The following sections assume that the former scenario is correct, based on the fact that prosimians do not possess an ortholog of the HBA gene that would have been the middle gene of the 3-gene set in the common ancestor of New World monkeys, Old World monkeys, and apes (= the HBA-T1 gene of Callicebus; fig. 7B). It should be noted that our inferences about the numbers of gene gains and losses are identical under both scenarios.
We analyzed genomic sequence data from 3 species of prosimian primates: the small-eared galago (*Otolemur garnettii*), the mouse lemur (*Microcebus murinus*), and the ring-tailed lemur (*Lemur catta*). All 3 species possess duplicated copies of HBA, and the mouse lemur and the galago also possess duplicated copies of HBZ (fig. 3). This suggests that the common ancestor of these 3 prosimian species had an \(\alpha\)-globin gene cluster with the following structure: 5' HBZ-T1, HBZ-T2, HBK, HBA-T1, HBA-T2, HBQ-3'. If this ancestral gene arrangement is correct, then single copies of HBZ and HBK have been secondarily lost in the ring-tailed lemur. Phylogeny reconstructions of flanking sequence indicate that the HBZ-T1 pseudogene of the mouse lemur is orthologous to the HBZ-T1 gene of the galago and that the HBZ-T2 genes in the galago and the mouse lemur have been secondarily lost in the ring-tailed lemur. Phylogeny reconstructions of flanking sequence indicate that the HBZ-T1 pseudogene of the mouse lemur is orthologous to the HBZ-T1 gene of the galago and that the HBZ-T2 genes in the galago and the mouse lemur are 1:1 orthologs as well (fig. 5). In contrast, the ring-tailed lemur has a single copy of HBZ. Whereas the 5' flanking sequence of this gene matches the 5' flanking sequence of HBZ-T2 in these other 2 species (fig. 5), this suggests that 1 HBZ gene was deleted from the \(\alpha\)-globin gene cluster of the ring-tailed lemur by an unequal crossing-over event similar to that shown in figure 4.

In the case of the HBA paralogs, all prosimians have 2 copies, one of which has become a pseudogene in the ring-tailed lemur by an unequal crossing-over event (fig. 3). Despite the fact that coding regions of the HBA paralogs have been homogenized by gene conversion, analyses of the flanking regions reveal that the HBA-T1 genes of all prosimians are 1:1 orthologs and likewise for the HBA-T2 copies (figs. 6 and 7B).

**Anthropoid Primates (Platyrrhini and Catarrhini)**

The \(\alpha\)-globin gene cluster of anthropoid primates appears to have undergone an especially high rate of turnover due to lineage-specific gains and losses of HBZ and HBA genes (fig. 7). Genomic sequence comparisons indicate that the \(\alpha\)-globin gene cluster in the ancestor of New World monkeys (platyrrhines) contained 2 copies of HBZ and 2 copies of HBA and that the \(\alpha\)-globin gene cluster in the common
ancestor of Old World monkeys and apes (catarrhines) contained 2 copies of HBZ and 3 copies of HBA. As in prosimians, phylogeny reconstructions based on coding sequence indicate that orthologous relationships among HBZ and HBA genes have been obscured by a history of concerted evolution, but analyses of flanking sequence enable us to resolve orthologous relationships in the majority of cases (figs. 5 and 6). Three of the platyrrhines in our data set, the titi monkey (*Callicebus moloch*), the owl monkey (*Aotus nancymaae*), and the marmoset (*Callithrix jacchus*), possess a single copy of HBZ, and the fourth species, the squirrel monkey (*Saimiri boliviensis*), has duplicate copies of HBZ. Comparisons of flanking sequence indicate that the *Saimiri* HBZ-T1 gene is orthologous to the HBZ-T1 gene of Old World monkeys and apes (catarrhines), whereas HBZ-T2 is orthologous to the 3' HBZ gene of catarrhines (figs. 5 and 7A). This indicates that an HBZ paralog was lost independently in each of the platyrhine species that have a single HBZ gene, and analysis of upstream and downstream flanking sequence indicates that these losses were due to unequal crossing-over, as in the ring-tailed lemur.

Phylogenetic analyses suggest that the presence of 3 copies of HBA in the ancestor of catarrhines was due to unequal crossing-over between chromosomes that originally possessed 2 copies of HBA. This conclusion is consistent with the observed distribution of homology blocks found in the α-globin gene cluster of apes (Shaw et al. 1991; Bailey et al. 1997). The upstream flanking sequence of HBA-T2 of most catarrhines is more closely related to the upstream sequence of the 3' HBZ gene of prosimians and platyrrhines. Conversely, the downstream sequence HBA-T2 of most catarrhines is more closely related to the downstream sequence of the 5' HBZ gene of prosimians and platyrrhines (fig. 6). These results suggest that the 5' HBA gene of most platyrhines is orthologous to the HBA-T2 gene of most catarrhines (figs. 6 and 7B). Additional phylogenetic analysis of a 2-kb alignment of flanking sequence upstream the start codon of the HBA genes of anthropoids (supplementary fig. S3, Supplementary Material Online) provided additional insights into the true set of orthologous relationships, although the presence of gene conversion tracts is also evident in the analyses of flanking sequences. Taken together, our analyses of upstream and downstream flanking sequence also suggest that, with the exception of the olive baboon (*Papio anubis*), the white-handed gibbon (*Hylobates klossi*), and the green monkey (*Chlorocebus aethiops*), all HBA-T2 genes of catarrhines are 1:1 orthologs (fig. 7B). The analysis of flanking sequences also revealed that the HBA-T2 gene...
of the dusky titi (C. moloch) originated via an independent, unequal crossing-over event. Finally, orthologs of the HBA-T1 gene of prosimians have been inactivated in all catarhines other than Chlorocebus, and they have been deleted independently in Chlorocebus and in all platyrhines. Accordingly, we infer that the ancestral HBA-T1 gene was deleted in the stem lineage of the platyrrhine clade and that the \( \alpha \)-globin gene cluster in the common ancestor of anthropoid primates had the following gene arrangement: 5' -HBZ-T1, HBZ-T2, HBK, HBA-T1, HBA-T2, HBA-T3, HBQ-3'.

The lineage-specific gains and losses of HBZ and HBA genes in primates mirror patterns that have been described in rodents, where unequal crossing-over events gave rise to functionally distinct copies of HBA in the deer mouse (Peromyscus maniculatus) and the Norway rat (R. norvegicus) (Storz, Hoffmann, et al. 2008). Although phylogeny reconstructions reveal that interparalog gene conversion is pervasive in coding regions, our analysis of flanking sequences revealed several instances where monophyly of paralogous sequences from the same species is attributable to lineage-specific gene duplications. In primates, for example, we identified 4 \( \alpha \)-like globin genes that were the products of de novo duplication events: HBZ-T2 in Pan, HBZ-T2 in Colobus, HBA-T2 in Callitrichus, and HBA-T2 in the common ancestor of platyrhines and catarhines. The latter gene has been secondarily lost several times independently, having been deleted in Callithrix, Hyllobates, and P. anubis and inactivated in Chlorocebus.

As an explanation for patterns of sequence similarity among paralogous genes within the same species, results of our comparative genomic analysis of the \( \alpha \)-globin gene family in mammals suggest that concerted evolution may not be as important as many previous workers had assumed. This conclusion is consistent with the results of several other comparative genomic studies of gene family evolution (Nei et al. 2000; Piontkivska et al. 2002; Eirin-Lopez et al. 2004; Nei and Rooney 2005; Rooney and Ward 2005).

Evolutionary Implications of Variation in Copy Number among Species

Results of our study reveal a high rate of differential gene gain and loss among the \( \alpha \)-globin gene clusters of different mammalian species. Over the course of mammalian evolution, we have documented the “birth” of new genes via duplication as well as “death” via inactivation or deletion. This “genomic revolving door” (Demuth et al. 2006) of gene gain and loss has resulted in continual turnover in the membership of the \( \alpha \)-globin gene family. Consequently, many mammals possess \( \alpha \)-like globin genes that have no orthologous counterparts in closely related species. In other cases, the ortholog of an apparently functional gene in one
species is a pseudogene in another species. For example, the ortholog of the HBZ-T3 gene in chimpanzee is a pseudogene in human, and the ortholog of the HBZ-T1 gene in humans is a pseudogene in chimpanzee (fig. 7A). Results of our detailed study of the z-globin gene family mirror the results of a genome-wide survey of size variation among mammalian gene families (Demuth et al. 2006). The analysis of Demuth et al. (2006) revealed that at least 6% of genes between human and chimpanzee are not orthologous. As these authors point out, this striking difference in gene content between the human and chimpanzee genomes stands in stark contrast to the well-documented 1.5% difference between orthologous nucleotide sequences.

The addition or subtraction of genes is expected to produce dosage imbalances that may often have deleterious effects. The adverse effects of such dosage imbalances have been well documented in the case of the adult z- and b-globin genes, as whole or partial gene deletions produce the thalassemia pathologies (Forget 2001; Higgs 2001; Lam and Jeffreys 2007). Although changes in gene dosage are generally expected to have deleterious effects, variation in gene copy number may also represent a source of potentially adaptive regulatory variation. It has been suggested that phenotypic differences among species are more commonly attributable to changes in gene regulation than changes in protein structure (King and Wilson 1975; Carroll 2005). The variation in globin gene copy number that we have documented among different mammalian lineages may constitute an important source of regulatory variation that affects physiologically important aspects of blood oxygen transport and aerobic energy metabolism.

Supplementary Material

Supplementary figures S1–S3 and tables S1 and S2 are available at Molecular Biology Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank K. Campbell, D. Irwin, A. Runck, and 4 anonymous reviewers for helpful comments and suggestions. This work was funded by a Postdoctoral Fellowship in Population Biology to F.G.H. from the University of Nebraska, an National Science Foundation grant to J.F.S. (DEB-0614342), a Layman Award to J.F.S., and an Interdisciplinary Research Grant to J.F.S. from the Nebraska Research Council.

Literature Cited


David Irwin, Associate Editor

Accepted January 1, 2008