

# Lineage-Specific Patterns of Functional Diversification in the $\alpha$ - and $\beta$ -Globin Gene Families of Tetrapod Vertebrates

Federico G. Hoffmann,<sup>1,2</sup> Jay F. Storz,<sup>2</sup> Thomas A. Gorr,<sup>3</sup> and Juan C. Opazo<sup>\*,4</sup>

<sup>1</sup>Instituto Carlos Chagas-ICC-Fiocruz, Curitiba, Brazil

<sup>2</sup>School of Biological Sciences, University of Nebraska, Lincoln

<sup>3</sup>Institute of Veterinary Physiology, Vetsuisse Faculty and Zurich Center for Integrative Human Physiology, University of Zurich, Zurich, Switzerland

<sup>4</sup>Instituto de Ecología y Evolución, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile

\*Corresponding author: E-mail: jopazo@gmail.com.

Associate editor: Scott Edwards

## Abstract

The  $\alpha$ - and  $\beta$ -globin gene families of jawed vertebrates have diversified with respect to both gene function and the developmental timing of gene expression. Phylogenetic reconstructions of globin gene family evolution have provided suggestive evidence that the developmental regulation of hemoglobin synthesis has evolved independently in multiple vertebrate lineages. For example, the embryonic  $\beta$ -like globin genes of birds and placental mammals are not 1:1 orthologs. Despite the similarity in developmental expression profiles, the genes are independently derived from lineage-specific duplications of a  $\beta$ -globin pro-ortholog. This suggests the possibility that other vertebrate taxa may also possess distinct repertoires of globin genes that were produced by repeated rounds of lineage-specific gene duplication and divergence. Until recently, investigations into this possibility have been hindered by the dearth of genomic sequence data from nonmammalian vertebrates. Here, we report new insights into globin gene family evolution that were provided by a phylogenetic analysis of vertebrate globins combined with a comparative genomic analysis of three key sauropsid taxa: a squamate reptile (anole lizard, *Anolis carolinensis*), a passeriform bird (zebra finch, *Taeniopygia guttata*), and a galliform bird (chicken, *Gallus gallus*). The main objectives of this study were 1) to characterize evolutionary changes in the size and membership composition of the  $\alpha$ - and  $\beta$ -globin gene families of tetrapod vertebrates and 2) to test whether functional diversification of the globin gene clusters occurred independently in different tetrapod lineages. Results of our comparative genomic analysis revealed several intriguing patterns of gene turnover in the globin gene clusters of different taxa. Lineage-specific differences in gene content were especially pronounced in the  $\beta$ -globin gene family, as phylogenetic reconstructions revealed that amphibians, lepidosaurs (as represented by anole lizard), archosaurs (as represented by zebra finch and chicken), and mammals each possess a distinct independently derived repertoire of  $\beta$ -like globin genes. In contrast to the ancient functional diversification of the  $\alpha$ -globin gene cluster in the stem lineage of tetrapods, the physiological division of labor between early- and late-expressed genes in the  $\beta$ -globin gene cluster appears to have evolved independently in several tetrapod lineages.

**Key words:** *Anolis*, gene duplication, gene family evolution, genome evolution, hemoglobin, zebra finch.

## Introduction

The functional and regulatory divergence of duplicated genes is known to play an extremely important role in evolutionary innovation (Ohno 1970; Zhang 2003; Taylor and Raes 2004; Lynch 2007). Repeated rounds of gene duplication and divergence can lead to the functional diversification of multigene families, whereby different gene family members acquire distinct biochemical functions and/or patterns of expression. The  $\alpha$ - and  $\beta$ -globin gene families of jawed vertebrates (gnathostomes) provide an excellent example of the kind of physiological versatility that can be attained through functional and regulatory divergence of duplicated genes that encode different subunit polypeptides of the same heteromeric protein. The hemoglobin of gnathostome vertebrates is a heterotetramer composed of two  $\alpha$ -chain subunits and two  $\beta$ -chain subunits that

are encoded by members of two paralogous gene families. The progenitors of the  $\alpha$ - and  $\beta$ -globin gene families were produced by tandem duplication of an ancestral single-copy hemoglobin gene  $\sim$ 450–500 Ma in the stem lineage of gnathostomes (Goodman et al. 1975, 1987; Czelusniak et al. 1982). The ancestral linkage arrangement of the proto  $\alpha$ - and  $\beta$ -globin genes has been retained in amphibians (Hentschel et al. 1979; Jeffreys et al. 1980; Kay et al. 1980; Hosbach et al. 1983; Fuchs et al. 2006) and teleost fish (Chan et al. 1997; Gillemans et al. 2003; Pisano et al. 2003). In birds and placental mammals, by contrast, the  $\alpha$ - and  $\beta$ -globin gene clusters are located on different chromosomes due to transposition of the proto  $\beta$ -globin gene to a new genomic location sometime after the stem lineage of amphibians split from the stem lineage of amniotes (Hardison 2008; Patel et al. 2008). In monotremes and marsupials, a vestige of the ancestral linkage arrangement is

present in the form of an orphaned  $\beta$ -like globin gene ( $\omega$ -globin) that is located at the 3' end of the  $\alpha$ -globin gene cluster (Wheeler et al. 2001, 2004; De Leo et al. 2005; Hoffmann et al. 2008b; Patel et al. 2008).

Over the course of vertebrate evolution, subsequent rounds of gene duplication and divergence gave rise to families of  $\alpha$ - and  $\beta$ -like globin genes that are ontogenetically regulated, such that functionally distinct hemoglobin isoforms are expressed in embryonic and adult erythroid cells (Collins and Weissman 1984; Hardison 1998, 2001). The  $\alpha$ - and  $\beta$ -like globin genes of *Xenopus* are differentially expressed in the larval and adult stages (Banville and Williams 1985a, 1985b; Fuchs et al. 2006), and the  $\alpha$ - and  $\beta$ -like globin genes of birds and mammals are also differentially expressed during prenatal development and postnatal life (Hardison 2001; Alev et al. 2008, 2009).

In the  $\alpha$ -globin gene family, the physiological division of labor between early- and late-expressed genes was established in the common ancestor of tetrapod vertebrates, and it appears to have been retained in all descendant lineages. The ancestral arrangement of the tetrapod  $\alpha$ -globin gene cluster consists of three linked genes, 5'- $\alpha^E$ ,  $\alpha^D$ , and  $\alpha^A$ -3' (Hoffmann and Storz 2007). The  $\alpha^E$ - and  $\alpha^A$ -globin genes originated via tandem duplication of an ancestral proto  $\alpha$ -globin gene after the stem lineage of fishes split from the stem lineage of tetrapods, and the  $\alpha^D$ -globin gene originated subsequently via tandem duplication of the  $\alpha^E$ -globin gene in the common ancestor of tetrapods (Hoffmann and Storz 2007). All tetrapods examined to date possess an ortholog of the early-expressed  $\alpha^E$ -globin gene at the 5' end of the  $\alpha$ -globin gene cluster and an ortholog of the late-expressed  $\alpha^A$ -globin gene at the 3' end of the cluster (Flint et al. 2001; Hughes et al. 2005; Cooper et al. 2006; Fuchs et al. 2006; Hoffmann and Storz 2007; Hoffmann et al. 2008b). In mammals, the  $\alpha^E$ - and  $\alpha^A$ -globin genes are often present in multiple copies, but the  $\alpha^D$ -globin gene has been deleted independently in multiple lineages (Hughes et al. 2005; Cooper et al. 2006; Hoffmann et al. 2008b). The  $\alpha^D$ -globin gene has also been secondarily lost from *Xenopus* (Fuchs et al. 2006; Hoffmann and Storz 2007), and it remains to be seen whether this gene is absent from the genomes of all amphibians. The  $\alpha^E$ -globin gene (known as  $\alpha^L$ -globin in amphibians,  $\pi$ -globin in birds, and  $\zeta$ -globin in mammals) is exclusively expressed in primitive erythroid cells derived from the yolk sac, and the adult  $\alpha^A$ -globin gene is expressed in definitive erythroid cells during later stages of prenatal development and postnatal life (Proudfoot et al. 1982; Higgs et al. 1989; Hardison 2001). The  $\alpha^A$ - and  $\alpha^D$ -globin genes are also present in squamate reptiles, and both genes exhibit elevated rates of amino acid substitution relative to their orthologous counterparts in turtles, tuatara, and birds (Gorr et al. 1998). In birds, the  $\alpha^D$ -globin gene is expressed in both embryonic and adult erythrocytes (Alev et al. 2008, 2009). By contrast, even in mammals that possess an intact, transcriptionally active copy of  $\alpha^D$ -globin, the product of this gene is not known to be assembled into functional hemoglobin tetramers (Goh et al. 2005).

In contrast to the ancient diversification of the  $\alpha$ -globin gene family, available evidence suggests that the physiological division of labor between early- and late-expressed  $\beta$ -like globin genes may have been repeatedly reinvented in different vertebrate lineages. For example, the embryonic " $\epsilon$ -globin" gene of chicken (*Gallus gallus*) is not a 1:1 ortholog of the embryonic  $\epsilon$ -globin gene of placental mammals as they were independently derived from lineage-specific duplications of a  $\beta$ -globin pro-ortholog (Goodman et al. 1987; Reitman et al. 1993; Aguilera et al. 2004; Patel et al. 2008; Opazo et al. 2008b). Indeed, even within mammals, early- and late-expressed  $\beta$ -like globin genes appear to have evolved independently in egg-laying monotremes and therians (placental mammals + marsupials; Opazo et al. 2008b). These patterns suggest that the developmental regulation of  $\beta$ -chain hemoglobin synthesis may have evolved multiple times independently. Until recently, investigations into this possibility were hindered by the fact that chicken was the only nonmammalian tetrapod for which a complete genome sequence was available. Fortunately, the availability of complete genome sequences from an amphibian (western clawed frog, *Xenopus tropicalis*), a squamate reptile (anole lizard, *Anolis carolinensis*), and a passeriform bird (zebra finch, *Taeniopygia guttata*) now permits a more comprehensive and detailed reconstruction of globin gene family evolution in tetrapod vertebrates. Here, we report new insights into globin gene family evolution that were provided by a phylogenetic analysis of vertebrate globins combined with a comparative analysis of genome sequences from three sauropsid taxa: anole lizard, zebra finch, and chicken. Comparative genomic analyses of the three sauropsid taxa can be expected to yield important insights into globin gene family evolution because lepidosaur reptiles (represented by anole lizard) diverged from the stem lineage of archosaurs (represented by birds) 275 Ma (Hedges 2009) and passeriform and galliform birds (represented by zebra finch and chicken, respectively) diverged  $\sim$ 100 Ma (Hedges et al. 2006).

The main objectives of this study were 1) to characterize evolutionary changes in the size and membership composition of the  $\alpha$ - and  $\beta$ -globin gene families of tetrapod vertebrates and 2) to test whether functional diversification of the globin gene clusters occurred independently in different tetrapod lineages. Results of our comparative genomic analysis revealed intriguing patterns of gene turnover in the orthologous globin gene clusters of different taxa and also revealed pronounced differences between the  $\alpha$ - and  $\beta$ -globin gene families in the pattern and timing of diversification. Specifically, our results revealed that the  $\beta$ -like globin genes diversified independently in several major lineages of tetrapod vertebrates: amphibians, lepidosaurs, archosaurs, and mammals.

## Materials and Methods

### Genomic Sequence Data and Annotation

We obtained genomic DNA sequences spanning the  $\alpha$ - and  $\beta$ -globin gene clusters of two bird species: zebra finch

(unpublished genome, the Genome Center at Washington University) ( $\alpha$ -globin gene cluster, chromosome 14: 2987641–2979228;  $\beta$ -globin gene cluster, chromosome 1B: 784737–915857) and chicken ( $\alpha$ -globin gene cluster, AY016020;  $\beta$ -globin gene cluster, L17432). We also searched the current assembly of the green anole lizard genome (unpublished genome, Broad Institute, release 54 of the ensembl database) for the presence of  $\alpha$ - and  $\beta$ -like globin genes. In all cases,  $\alpha$ - and  $\beta$ -like globin genes were annotated by comparison with other globin genes that were already present in the database. We also used the program Genscan (Burge and Karlin 1997) to identify and annotate additional genes lying upstream and downstream of the globin gene clusters. In most cases, coding sequences derived from the genome assembly were validated by comparison with the corresponding EST database.

In addition to the genome sequences from anole lizard, zebra finch, and chicken, we also obtained sequences from a phylogenetically diverse collection of other vertebrate species in order to reconstruct pathways of globin gene family evolution during the radiation of tetrapods. In the case of the  $\alpha$ -like globin genes, we annotated sequences of  $\alpha$ -like globin genes from human (*Homo sapiens* [NG\_000006]) and platypus (*Ornithorhynchus anatinus* [AC203513]), and we obtained sequences of  $\alpha$ -like globin genes from frog (*X. tropicalis* [NM\_001005092 and NM\_203529]), salamander (*Pleurodeles waltl* [M13365 and X14226]), yellow-foot tortoise (*Geochelone denticulata* [AF499739]), and zebrafish (*Danio rerio* [NM\_131257]). In the case of the  $\beta$ -like globin genes, we obtained genomic sequences spanning the  $\beta$ -globin gene cluster of two placental mammals (human, *H. sapiens* [AC104389] and cat, *Felis catus* [AC129072]), one amphibian (western clawed frog, *X. tropicalis* [scaffold 733: 145910–445992]), and four fishes (zebrafish, *D. rerio* [chromosome 3: 52848044–52988682]; medaka, *Oryzias latipes* [chromosome 19: 1414512–1555137]; fugu, *Takifugu rubripes* [chromosome Un: 14351875–14492550]; and pufferfish, *Tetraodon nigroviridis* [chromosome 2: 3478461–3619109]). For each species, we identified globin genes in the unannotated genomic sequence by using the program Genscan (Burge and Karlin 1997) and by comparing known exon sequences to genomic contigs using the program BLAST2, version 2.2 (Tatusova and Madden 1999).

In order to obtain a more detailed view of the evolutionary history of the  $\beta$ -like globin genes in sauropsid vertebrates, we also assembled a data set of amino acid sequences that included  $\beta$ -globin sequences from a diverse array of reptiles. For comparative purposes, this data set included conceptual translations of all functional  $\beta$ -like globin genes from anole lizard, chicken, zebra finch, human, and cat, as well as amino acid sequences from an additional squamate reptile (common iguana, *Iguana iguana*, [P18987]), a sphenodont reptile (tuatara, *Sphenodon punctatus* [P10060 and P10061]), crocodylians (Nile crocodile, *Crocodylus niloticus* [P02129], and American alligator, *Alligator mississippiensis* [P02130]), and testudines (loggerhead turtle, *Caretta caretta* [Q10733], and Galapagos tortoise, G.

*nigra* [P83123]). We also used new amino acid sequences from paralogous  $\beta$ I- and  $\beta$ II-globins of the Indian python, *Python molurus* (Squamata: Boidae; Stoeckelhuber 1992; Gorr 1993) and the  $\beta$ II-globin paralog of the common iguana (see below). Two  $\beta$ -globin paralogs from the western clawed frog were included as outgroup sequences.

### Amino Acid Sequencing

We determined the complete primary sequence of the  $\beta$ II-globin paralog of the common iguana (*I. iguana*) and the Galapagos marine iguana (*Amblyrhynchus cristatus*) by means of automated Edman degradation in conjunction with amino acid analyses of peptides separated by high-performance liquid chromatography (HPLC). Specifically, we conducted N-terminal sequencing of the isolated and lyophilized native chains (Rücknagel et al. 1988) together with gas-phase sequencing of HPLC-purified tryptic peptides from reduced pyridine-ethylated chains (Gorr 1993; Stoeckelhuber et al. 2002). This latter reaction uses vinyl pyridine to alkylate reactive cyteiny SH-groups in order to irreversibly block disulfide-based aggregates of  $\beta$ -chain subunits. The  $\beta$ II-globin amino acid sequences from the common iguana and the marine iguana were deposited in the UniProt Knowledgebase under accession nos. P86390 and P86391.

### Phylogeny of $\alpha$ - and $\beta$ -Like Globin Genes

Sequences of vertebrate globin genes were aligned using the program MUSCLE (Edgar 2004), as implemented in the European Bioinformatics Institute Web server (<http://www.ebi.ac.uk>). Phylogenetic relationships were inferred using Bayesian and maximum likelihood approaches. Bayesian analyses were conducted in Mr. Bayes v.3.1.2 (Ronquist and Huelsenbeck 2003) running six simultaneous chains for 5,000,000 generations, sampling every 1,000 generations, and using default priors. Support for the nodes and parameter estimates were derived from a majority rule consensus of the last 2,500 trees. The average standard deviation of split frequencies remained less than 0.01 after the burn-in threshold. Maximum likelihood analyses were conducted in Treefinder, version October 2008 (Jobb et al. 2004), and support for the nodes was evaluated with 1,000 bootstrap pseudoreplicates. In both Bayesian and maximum likelihood analyses, different models of nucleotide substitution were permitted for each codon position. We used the “propose model” tool of Treefinder to select the best fitting model of nucleotide substitution for each codon position in the maximum likelihood analysis. Model selection was based on the Akaike information criterion with correction for small sample size. In the case of the amino acid data set, the phylogeny reconstruction was conducted with Mr. Bayes v.3.1.2 (Ronquist and Huelsenbeck 2003) using a JTT model of amino acid substitution, running eight simultaneous chains for 5,000,000 generations, sampling every 1,000 generations, and using default priors. Support for the nodes and parameter estimates were derived from a majority rule consensus of the last 2,500 trees. The average standard deviation of split frequencies

remained less than 0.01 after the burn-in threshold. For the phylogenetic analysis of  $\beta$ -globin amino acid sequences, we only included sequences from reptile taxa for which we successfully recovered all expressed adult  $\beta$ -globins. For this reason, we excluded marine iguana from the analysis because we successfully sequenced only one of the two adult  $\beta$ -globin paralogs from this species.

### Inferring Orthologous Relationships

To examine the genomic structure of the  $\alpha$ - and  $\beta$ -globin gene clusters of zebra finch, we conducted pairwise comparisons of sequence similarity with the orthologous chicken gene cluster using the Pipmaker program (Schwartz et al. 2000). Because interparalog gene conversion is typically restricted to the coding regions of globin genes, we inferred orthologous relationships among  $\beta$ -like globin genes by reconstructing phylogenetic relationships of flanking sequence and intron 2 sequence (Hardison and Gelinas 1986; Hardison and Miller 1993; Storz et al. 2007; Hoffmann et al. 2008a, 2008b; Opazo et al. 2008a, 2008b; Opazo et al. 2009; Runck et al. 2009). We conducted phylogeny reconstructions on three nonoverlapping sequence alignments: 1 kb of flanking sequence immediately upstream of the initiation codon, 2385 bp of intron 2 sequence, and 1 kb of flanking sequence immediately downstream of the termination codon. Because previous studies have demonstrated that ectopic recombinational exchanges are restricted to the 5' ends of chicken  $\beta$ -like globin genes (including exons 1 and 2) and upstream flanking regions (Dodgson et al. 1983; Reitman et al. 1993), sequence variation in intron 2 and the downstream flanking region should be especially useful for the purpose of assigning orthologous relationships. Phylogenetic analyses of each data partition were conducted in a maximum likelihood framework as described above.

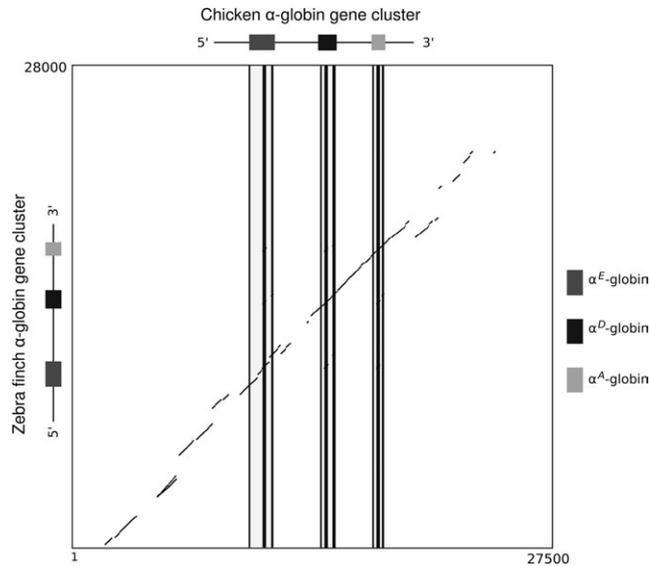
## Results

### Genomic Structure and Evolution of the Saurosid $\alpha$ -Globin Gene Cluster

The two bird species included in our study, zebra finch (Passeriformes: Estrildidae) and chicken (Galliformes: Phasianidae), have both retained the ancestral complement of developmentally regulated  $\alpha$ -like globin genes: 5'- $\alpha^E$ ,  $\alpha^D$ , and  $\alpha^A$ -3'. Pairwise dot plots between the  $\alpha$ -globin gene clusters of these two species revealed high levels of sequence conservation across both coding and noncoding regions (fig. 1). Phylogenies based on coding sequence indicate that the three  $\alpha$ -like globin genes in the zebra finch cluster are 1:1 orthologs of the three  $\alpha$ -like genes in the chicken cluster (fig. 2). The  $\alpha$ -globin gene clusters of both species also retained very similar patterns of intergenic spacing: From the initiation codon of  $\alpha^E$ -globin to the termination codon of  $\alpha^A$ -globin, the gene cluster spanned 8,414 bp in zebra finch and 7,379 bp in chicken. In both species, the  $\alpha$ -globin gene cluster was flanked by an ortholog of chromosome 16 open reading frame 35 (C16Orf35) on the 5' side and by an ortholog of the transmembrane protein 6 on the 3' side.

In the anole lizard genome, we identified two putatively functional  $\alpha$ -like globin genes that were located in different scaffolds (2790: 1–15038 and 1188: 1–163036). Phylogenies based on coding sequence indicate that one of the genes is orthologous to the  $\alpha^A$ -globin gene of other tetrapods, and the other gene is orthologous to the  $\alpha^D$ -globin gene of other amniotes (fig. 2). We did not find an ortholog of the embryonic  $\alpha^E$ -globin gene in the anole lizard genome. The  $\alpha^D$ -globin gene of this species is located on scaffold 2790, and its position is syntenic with the  $\alpha$ -globin gene cluster of other vertebrates: On the 5' end, the gene is flanked by an ortholog of C16Orf35, and on the 3' end, it is flanked by a highly distinct globin gene that exhibited a strong match to the recently discovered “globin Y” (GbY) gene. Bioinformatic analyses reported by Fuchs et al. (2006) indicated that the GbY gene is most closely related to cytoglobin, a respiratory globin protein that is expressed in fibroblast-related cells and distinct neuronal cells (Burmester et al. 2004; Hankeln et al. 2004; Nakatani et al. 2004; Schmidt et al. 2004; Hankeln et al. 2005; Schmidt et al. 2005). To date, the GbY gene has only been characterized in *Xenopus* (Fuchs et al. 2006) and platypus (Patel et al. 2008). The apparently orthologous GbY genes of *Xenopus*, anole lizard, and platypus are located in syntenic positions at the 3' end of the  $\alpha$ -globin gene cluster. The anole lizard GbY gene exhibits the typical 3-exon/2-intron globin gene structure, and a conceptual translation of the coding sequence yields a polypeptide of 151 amino acids that shows 35% identity/60% similarity to the GbY sequences from *Xenopus* and platypus and 28% identity/50% similarity to tetrapod cytoglobins. Despite its close linkage to the  $\alpha$ -like globin genes, the GbY gene does not encode a subunit polypeptide of hemoglobin and its function remains a mystery (Fuchs et al. 2006).

Interestingly, the *Anolis* ortholog of the adult  $\alpha^A$ -globin gene is located on scaffold 1188, and it is flanked by an ortholog of the chicken adenylate cyclase 9 gene (ADCY9; located on chromosome 14: 13294809–13353192) on the 5' end, and an ortholog of the chicken GSG1L gene (located on chromosome 14: 12411725–1274104) on the 3' end. The chicken ADCY9 and GSG1L genes are located on opposite sides of the chicken  $\alpha$ -globin cluster, which is located on chromosome 14: 12722493–12730496. This seemingly anomalous arrangement of  $\alpha$ -like globins in anole lizard is probably attributable to an error in the current genome assembly. If the position of the lizard  $\alpha^A$ -globin gene is not an artifact of the current assembly, it would represent a major departure from the canonical gene arrangement of all other tetrapod vertebrates studied to date, where the  $\alpha$ -like globin genes are always flanked by the genes N-methylpurine-DNA glycosylase and C16Orf35 at the 5' end of the gene cluster (Flint et al. 2001; Hughes et al. 2005; Hardison 2008; Patel et al. 2008; Hoffmann et al. 2008b). An alignment of  $\alpha$ -globin amino acid sequences from *Anolis*, chicken, and zebra finch is provided in [supplementary figure S1A](#) (Supplementary Material online).

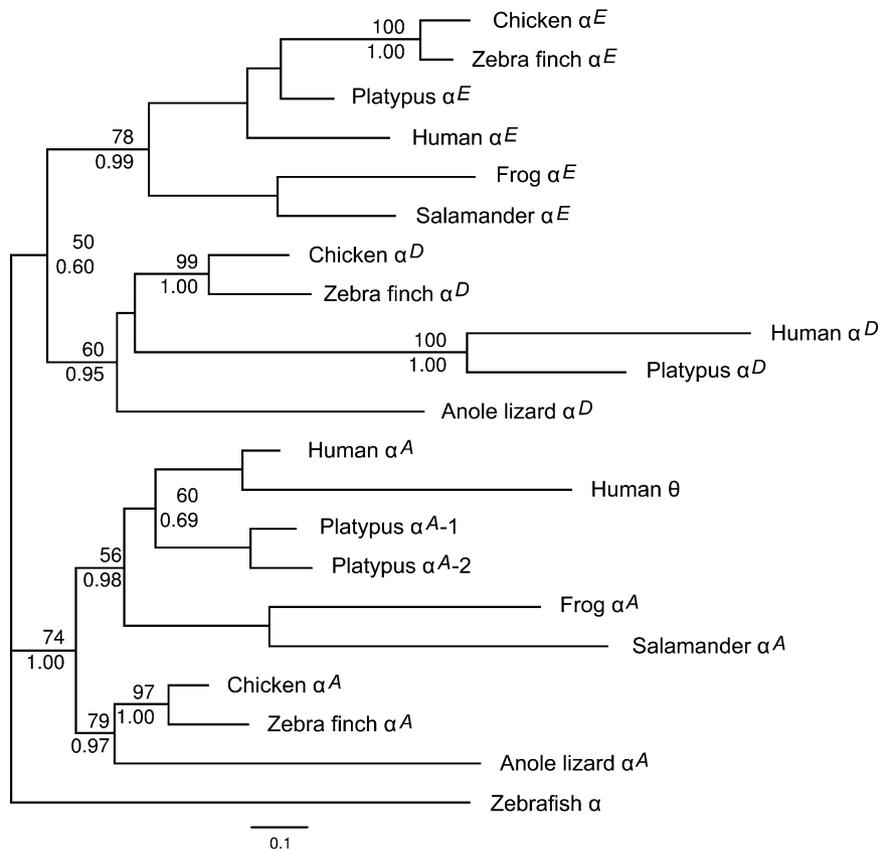


**Fig. 1.** Dot plot of sequence similarity between the  $\alpha$ -globin gene clusters of chicken (*Gallus gallus*) and zebra finch (*Taeniopygia guttata*).

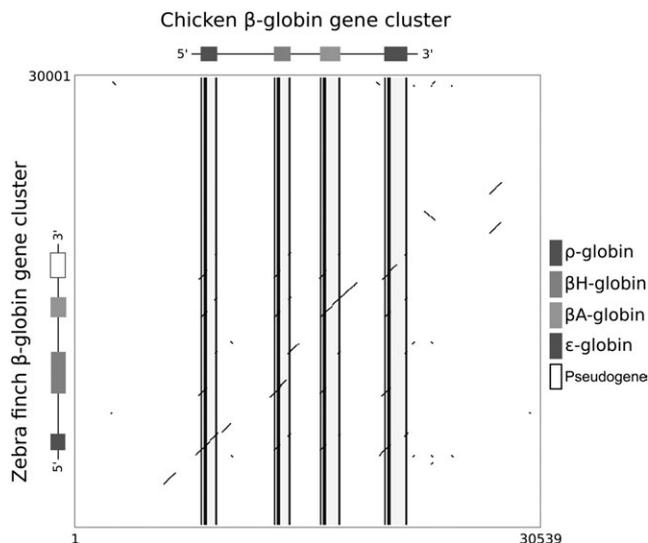
### Genomic Structure and Evolution of the Sauropsid $\beta$ -Globin Gene Cluster

In the zebra finch, the  $\beta$ -globin gene cluster contains four genes (from 5' to 3'):  $\rho$ -globin,  $\beta^H$ -globin,  $\beta^A$ -globin, and

an  $\varepsilon$ -globin pseudogene-3' (fig. 3). As with the  $\alpha$ -globin gene cluster, the  $\beta$ -globin gene clusters of zebra finch and chicken have remained very similar in terms of intergenic spacing: From the initiation codon of  $\rho$ -globin to the



**Fig. 2.** Maximum likelihood phylogram depicting relationships among the  $\alpha$ -globin sequences of vertebrates based on nucleotide sequence data. The sequence of the zebrafish (*Danio rerio*) was used as outgroup. Values on the nodes denote bootstrap support values (above) and Bayesian posterior probabilities (below).



**Fig. 3.** Dot plot of sequence similarity between the  $\beta$ -globin gene clusters of chicken (*Gallus gallus*) and zebra finch (*Taeniopygia guttata*).

termination codon of  $\epsilon$ -globin, the zebra finch cluster spans 13,231 bp and the chicken cluster spans 13,691 bp. The zebra finch  $\beta$ -globin gene cluster is flanked on both sides by multiple odorant receptors genes, as is the case with all other amniotes studied to date (Bulger et al. 1999; Hardison 2008; Patel et al. 2008). Pairwise sequence comparison between zebra finch and chicken revealed that high levels of sequence conservation were largely restricted to coding regions and proximal flanking regions (fig. 3).

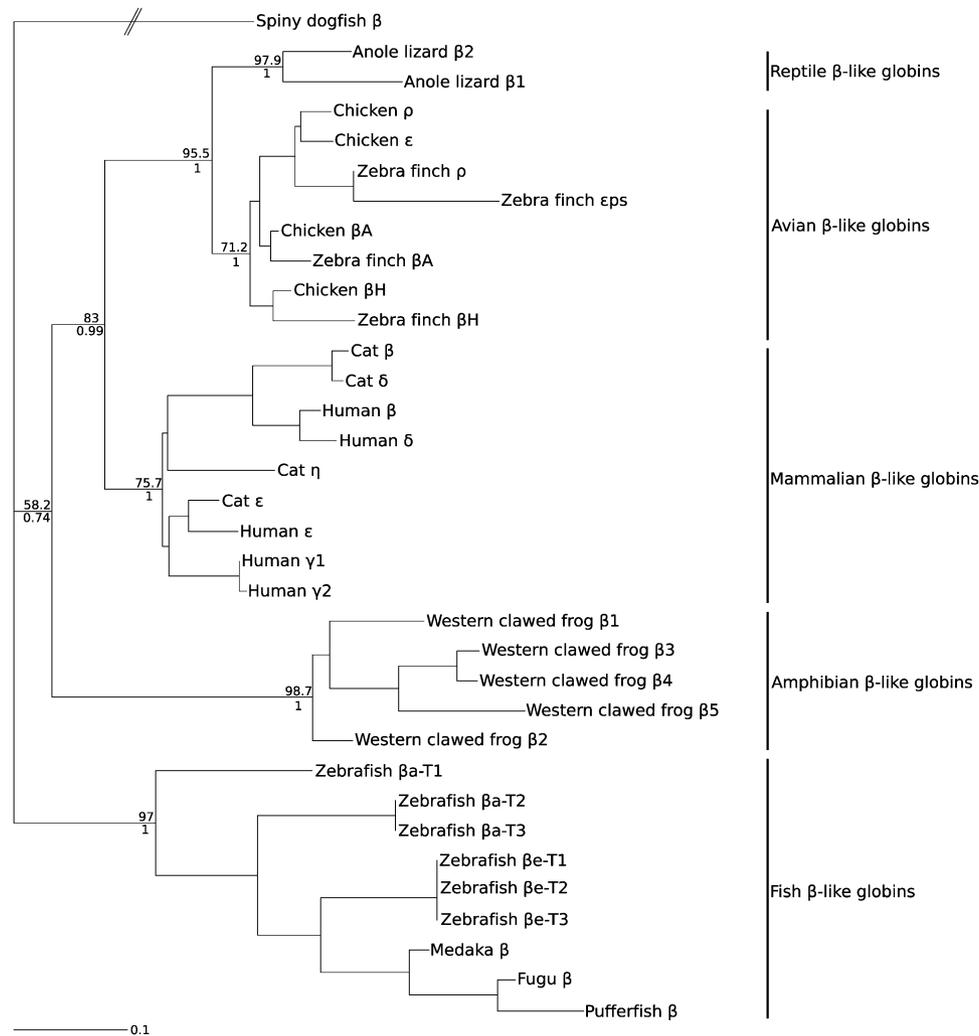
The annotation of  $\beta$ -like globin genes in zebra finch revealed that two of the four predicted  $\beta$ -like globin genes,  $\beta^A$ - and  $\epsilon$ -globin, contained insertions and deletions in the coding sequence that would render them nonfunctional. The  $\beta^A$ -globin sequence contained two single nucleotide insertions, and the  $\epsilon$ -globin sequence contained multiple frame-shifting indels. In comparison with the other early-expressed gene,  $\rho$ -globin, the putative  $\epsilon$ -globin pseudogene contained four indels in exon 2, ranging 6–50 bp in length, and three indels in exon 3, ranging 1–17 bp in length. To verify the authenticity of these inactivating indels in the zebra finch  $\beta^A$ - and  $\epsilon$ -globin genes, we searched the corresponding EST database and found a sequence corresponding to the  $\beta^A$ -globin gene that had a perfectly intact open reading frame. We therefore used the sequence from the EST record for all phylogenetic analyses. We did not find an EST record that matched the  $\epsilon$ -globin sequence, which suggests that this gene is not expressed in zebra finch.

In the anole lizard genome, we found two putatively functional  $\beta$ -like globin genes that were located in different scaffolds (7008: 1–4809 and 3777: 1–10310), and Genscan searches did not reveal the presence of any additional genes upstream or downstream of the  $\beta$ -like globin genes in either scaffold. Thus, the current state of the anole lizard genome assembly does not permit inferences regarding patterns of conserved synteny in the  $\beta$ -globin gene cluster. The two  $\beta$ -globin paralogs of *Anolis* were distinguished by

a total of 36 amino acid substitutions (24.49% sequence divergence). An alignment of  $\beta$ -globin amino acid sequences from *Anolis*, chicken, and zebra finch is provided in [supplementary figure S1B](#) (Supplementary Material online).

### Inferring Orthologous Relationships among $\beta$ -Like Globin Genes

If the functional diversification of the  $\beta$ -globin gene family predated the radiation of tetrapod vertebrates, then we should recover a phylogenetic tree with the following topological features: 1) The early-expressed genes of all taxa group together in one clade of orthologous sequences, 2) the late-expressed genes from these same taxa group together in a sister clade of orthologous sequences, and 3) sequences within each of the two paralogous sister clades independently recover the known organismal phylogeny: (amphibians (mammals (lepidosaurs, archosaurs))). Contrary to these expectations, maximum likelihood and Bayesian phylogeny reconstructions revealed that  $\beta$ -like globin genes of amphibians, reptiles, birds, and mammals were all reciprocally monophyletic relative to one another (fig. 4). This phylogenetic pattern indicates that the  $\beta$ -globin gene clusters diversified independently in each of these different tetrapod lineages. Each of the main groups of tetrapods inherited an ortholog of the same proto  $\beta$ -globin gene, which then underwent one or more rounds of duplication and divergence to produce distinct repertoires of  $\beta$ -like globin genes in each descendant lineage. These lineage-specific patterns had been previously documented in the case of birds (as represented by chicken alone) and mammals (Patel et al. 2008; Opazo et al. 2008a, 2008b). The addition of genomic sequences from *Xenopus* and anole lizard revealed that amphibians and squamate reptiles have also invented their own distinct repertoires of  $\beta$ -like globin genes. The addition of genomic sequence from zebra finch revealed that passeriform and galliform birds have retained the same ancestral complement of



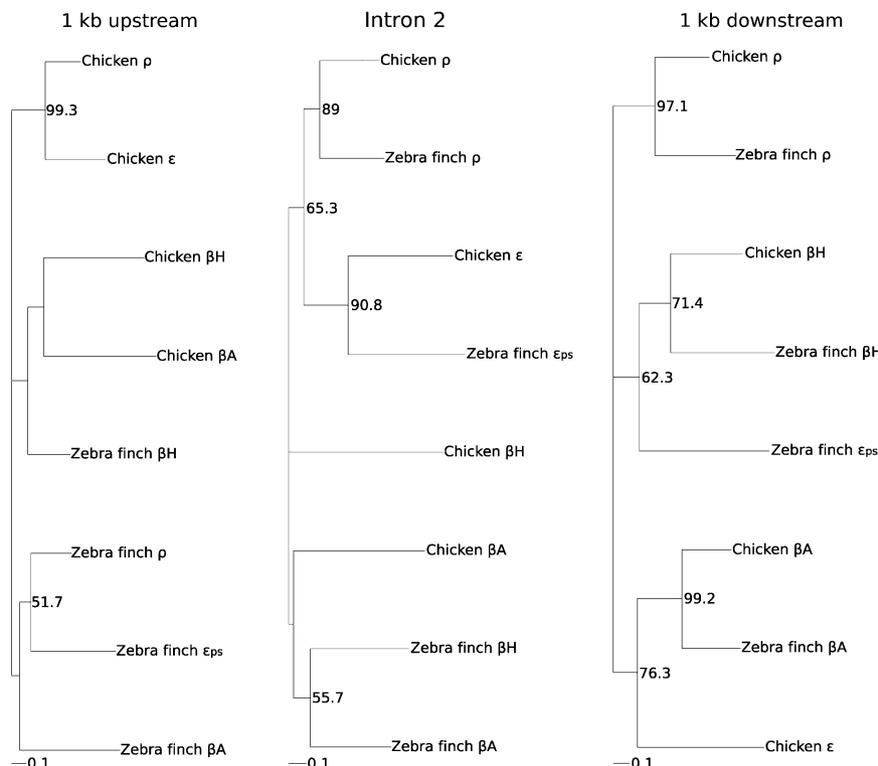
**Fig. 4.** Maximum likelihood phylogram depicting relationships among the  $\beta$ -globin sequences of vertebrates based on nucleotide sequence data. The sequence of the spiny dogfish (*Squalus acanthias*) was used as outgroup. Values on the nodes denote bootstrap support values (above) and Bayesian posterior probabilities (below).

$\beta$ -like globin genes. This was not a foregone conclusion because the estimated date of the duplication event that gave rise to the  $\rho$ -/ $\epsilon$ -globin gene pair in chicken was reported to be 49.4–55.1 Ma (Aguileta et al. 2006), whereas the estimated date of divergence between passeriformes and galiliformes is thought to be  $\sim$ 100 Ma (Hedges et al. 2006).

Within birds, phylogenies based on coding sequence were able to resolve orthologous relationships for the adult  $\beta$ -globin genes but not the early-expressed genes (fig. 5). Phylogeny reconstructions based on coding sequence and upstream flanking sequence depict a clear history of concerted evolution between the two early-expressed genes, as paralogous  $\rho$ - and  $\epsilon$ -globin sequences from the same species group together in the tree to the exclusion of orthologous sequences in other species (fig. 5). Pairwise alignment of the zebra finch  $\rho$ - and  $\epsilon$ -globin sequences shows an increased sequence similarity in the upstream sequence that is attributable to  $\rho \rightarrow \epsilon$  gene conversion. The conversion tract was 433 bp long, starting 90 bp upstream of the start codon and spanning exon 2 of the  $\epsilon$ -globin gene

(Dodgson et al. 1983). Because intron 2 and downstream flanking sequence have not experienced a history of gene conversion, phylogeny reconstructions of these noncoding regions can be expected to recover the true history of gene duplication and species divergence. Accordingly, phylogeny reconstructions based on intron 2 and downstream flanking sequence conclusively demonstrate that the  $\rho$ -,  $\beta^H$ -,  $\beta^A$ -, and  $\epsilon$ -globin genes of zebra finch and chicken are 1:1 orthologs (fig. 5).

The phylogeny reconstruction based on amino acid sequences revealed that reptilian  $\beta$ -globins form a single monophyletic group, although support for the corresponding node was relatively weak (fig. 6). In the tree based on amino acid sequences, archosaur  $\beta$ -globins grouped together in a clade that was sister to the  $\beta^D$ -globin paralog of the tuatara, and lepidosaur  $\beta$ -globins grouped together in a clade that also included the  $\beta^2$ -globin paralog of the Indian python and the  $\beta^A$ -globin paralog of the tuatara. In addition, the two testudine  $\beta$ -globins form a monophyletic group and the  $\beta^1$ -globin paralog of the Indian python



**FIG. 5.** Maximum likelihood phylogram depicting relationships among  $\beta$ -like globin genes in chicken and zebra finch based on 1 kb of the 5' flanking sequence (left), intron 2 (center), and 1 kb of 3' flanking sequence (right). Measures for support for the relevant nodes are presented as bootstrap values.

represents the most basal lineage. This phylogeny reconstruction indicates that the  $\beta_1$  globin genes of anole lizard and common iguana are 1:1 orthologs, as are the  $\beta_2$  globin genes from the same species pair (fig. 6). The  $\beta_1$  and  $\beta_2$  globin genes of anole lizard and iguana appear to be products of a lizard-specific duplication event (and the two iguana  $\beta$ -globin sequences correspond to the " $\beta_{11a}$ " and " $\beta_{11b}$ " sequences in Gorr et al. 1998). The phylogenetic positions of  $\beta$ -globin paralogs from tuatara and Indian python suggest that the common ancestor of sauropsids may have possessed a fairly diverse repertoire of  $\beta$ -like globin genes and that the differential loss and retention of these genes may account for much of the observed variation in gene content among contemporary reptiles. Due to the limited resolution afforded by amino acid sequences, we will ultimately need genomic sequence data from turtles, tuatara, and snakes to reconstruct pathways of  $\beta$ -globin gene family evolution during the basal radiation of nonavian reptiles. We also reconstructed phylogenetic relationships using a mixed model approach. The estimated topology showed two minor discrepancies relative to the original tree. The first discrepancy concerns the outgroup sequences, where relationships among the mammalian  $\epsilon$ -,  $\gamma$ -, and  $\eta$ -globin genes were not resolved. The second discrepancy is that the chicken and zebra finch  $\beta^A$ -globin genes were not grouped together, and instead the zebra finch  $\beta^A$ -globin gene was placed sister to the crocodilian  $\beta$ -globin genes. This latter result is clearly incorrect because trees based on DNA sequence variation in intron 2 and the 3' flanking

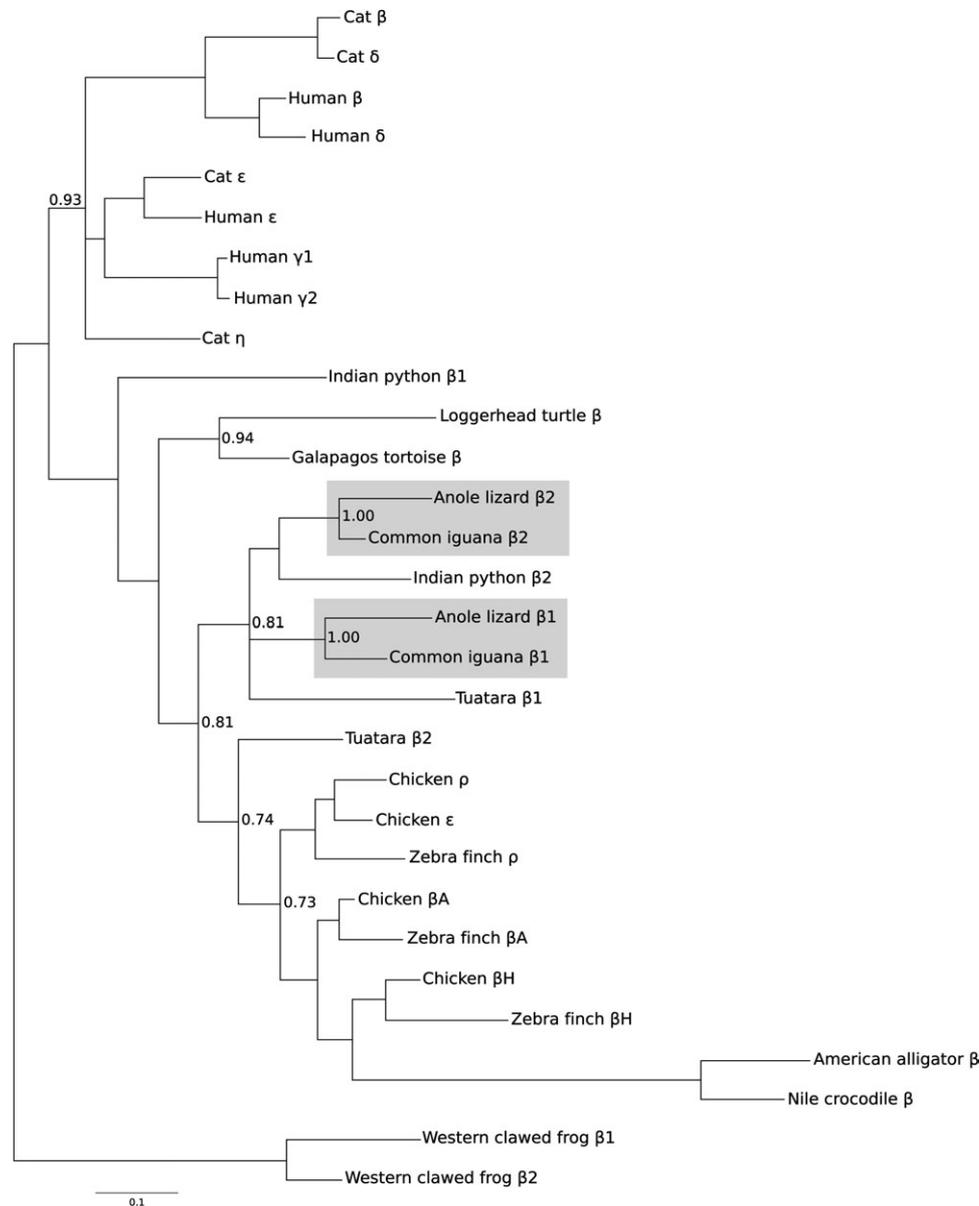
region also demonstrate that the chicken and zebra finch  $\beta^A$ -globin genes are 1:1 orthologs (fig. 5).

## Discussion

In recent years, comparative analyses of completely sequenced genomes have revealed extensive variation in the size and membership composition of multigene families among species (Demuth and Hahn 2009). This variation results from differential gene gains via duplication and gene losses via deletion or inactivation. In many cases, the resultant changes in gene family composition may account for important phenotypic differences between species. Our comparative genomic analysis of the  $\alpha$ - and  $\beta$ -globin gene clusters revealed a number of significant differences in gene content among vertebrate taxa. These differences were especially pronounced in the case of the  $\beta$ -globin gene cluster, as our phylogenetic reconstructions revealed that amphibians, lepidosaurs (as represented by anole lizard), archosaurs (as represented by zebra finch and chicken), and mammals each possess a distinct independently derived repertoire of  $\beta$ -like globin genes.

## Variation in Gene Family Size and Membership Composition among Taxa

In the case of the  $\alpha$ -globin gene family, the anole lizard emerged as the only gnathostome vertebrate species yet examined that does not possess an ortholog of the embryonic  $\alpha^E$ -globin gene. Given that the  $\alpha^D$ -globin gene



**Fig. 6.** Bayesian phylogram depicting relationships among the  $\beta$ -like globin genes of amniote vertebrates based on amino acid sequence data. Two  $\beta$ -globin sequences from western clawed frog (*Xenopus tropicalis*) were used as outgroups. Values on the nodes correspond to Bayesian posterior probabilities.

originated via duplication of a proto  $\alpha^E$ -globin gene that had an ancestral larval/embryonic function (Hoffmann and Storz 2007), it may be that  $\alpha^D$ -containing hemoglobin (*HbD*) performs the necessary oxygen transport functions during early stages of embryonic development in lizards and other reptiles that do not possess an ortholog of  $\alpha^E$ -globin. Developmental studies of hemoglobin synthesis will be necessary to test this hypothesis about the interchangeability of  $\alpha^E$ - and  $\alpha^D$ -globin in reptiles.

Although zebra finch and chicken possess an identical complement of  $\alpha^E$ -,  $\alpha^D$ -, and  $\alpha^A$ -globin genes, the  $\alpha^D$ -globin gene exhibits very different patterns of prenatal expression in these two species (Alev et al. 2009). In primitive erythroid cells of zebra finch embryos, the rank order of transcript abundance is  $\alpha^D > \alpha^E > \alpha^A$ , and in erythroid cells of

chicken at the same stage of development, the rank order of transcript abundance is  $\alpha^E > \alpha^A > \alpha^D$  (Alev et al. 2008, 2009). In the definitive erythroid cells of adult birds, the *HbD* isoform typically constitutes the minor fraction of adult hemoglobin and  $\alpha^A$ -containing hemoglobin typically constitutes the major fraction, although the relative abundance of the two isoforms is quite variable among species (Borgese and Bertles 1965; Brown and Ingram 1974; Hiebl et al. 1987; Ikehara et al. 1997). It will be important to achieve a more complete understanding of the physiological division of labor among the  $\alpha^E$ -,  $\alpha^D$ -, and  $\alpha^A$ -containing hemoglobin isoforms during different stages of development in birds and nonavian reptiles.

In the case of the  $\beta$ -globin gene family, the one notable difference between the two birds included in our analysis is

that the ortholog of the chicken  $\varepsilon$ -globin gene is a pseudogene in zebra finch. This may seem surprising because  $\varepsilon$ -globin is the most highly expressed  $\beta$ -like globin gene during early stages of chicken embryogenesis (Alev et al. 2008), and early-expressed globin genes are typically subject to especially stringent levels of functional constraint (Aguileta et al. 2004). However, the  $\rho$ - and  $\varepsilon$ -globin genes are very similar in sequence due to a history of interparalog gene conversion (supplementary fig. S1, Supplementary Material online). Concerted evolution between the two early-expressed paralogs may have produced a degree of functional redundancy that would have helped mitigate the effects of inactivating the  $\varepsilon$ -globin gene. Similarly, differential loss and retention of embryonic  $\beta$ -like globin genes during the basal radiation of placental mammals have produced extensive variation in the complement of early-expressed genes in contemporary species (Opazo et al. 2008a).

Results of our comparisons between zebra finch and chicken indicate that the size and membership composition of the  $\alpha$ - and  $\beta$ -globin gene families are far more stable in birds than in mammals. Zebra finch (Passeriformes) and chicken (Galliformes) represent two highly distinct avian lineages that are thought to have diverged  $\sim 100$  Ma (Hedges et al. 2006). Despite this ancient divergence, zebra finch and chicken have each retained an identical complement of three orthologous  $\alpha$ -like globin genes. By contrast, over a similar timescale of divergence, gene turnover in the mammalian  $\alpha$ -globin gene family has produced extensive variation in gene copy number among species (Hoffmann et al. 2008b). For example, guinea pig and rat, which diverged  $\sim 60$  Ma (Steppan et al. 2004), possess three and seven functional  $\alpha$ -like globin genes, respectively (Storz et al. 2008). Zebra finch and chicken have also retained an identical complement of orthologous  $\beta$ -like globin genes, with the exception that  $\varepsilon$ -globin has been inactivated in zebra finch. Again, the stability of the avian  $\beta$ -globin gene cluster stands in marked contrast to the pattern of gene turnover that has been documented in mammals (Opazo et al. 2008a, 2008b, 2009). Within rodents, for example, guinea pig and rat possess two and eight functional  $\beta$ -like globin genes, respectively (Hoffmann et al. 2008a).

In addition to the pronounced differences in genomic stability, the globin gene clusters of birds and mammals also differ with respect to the association between linear gene order and the temporal order of expression during development. Genes in the mammalian  $\alpha$ - and  $\beta$ -globin gene clusters are almost always colinear with their temporal order of expression: The early-expressed genes are located at the 5' end of the cluster and the later-expressed genes are located at the 3' end. The only known exceptions to this pattern are attributable to en bloc duplications of multiple genes, where early-expressed genes in the 3' duplication block are located downstream of late-expressed genes in the 5' duplication block (Townes et al. 1984; Cheng et al. 1987; Hoffmann et al. 2008a). In the chicken  $\beta$ -globin gene cluster, by contrast, the embryonic  $\rho$ - and  $\varepsilon$ -globin genes demarcate the 5' and 3' ends of the gene

cluster, respectively, and the late-expressed  $\beta^H$ - and  $\beta^A$ -globin genes are sandwiched in between (Bruns and Ingram 1973a, 1973b; Reitman et al. 1993).

### Different Timescales of Functional Diversification in the $\alpha$ - and $\beta$ -Globin Gene Families

Results of our phylogenetic analysis confirm that the  $\alpha$ - and  $\beta$ -like globin gene families of tetrapods have diversified over very different timescales. In the  $\alpha$ -globin gene cluster, a division of labor between early- and late-expressed genes was established in the common ancestor of tetrapods, and all descendant lineages appear to have retained the same basic mechanism for the developmental regulation of  $\alpha$ -chain hemoglobin synthesis: The  $\alpha$ -chain subunits of embryonic hemoglobin are encoded by  $\alpha^E$ -globin and the  $\alpha$ -chain products of adult hemoglobin are encoded by  $\alpha^A$ -globin. The anole lizard represents the one notable exception to this pattern due to the apparent absence of  $\alpha^E$ -globin in this species.

In contrast to the ancient functional diversification of the  $\alpha$ -globin gene cluster, the physiological division of labor between early- and late-expressed genes in the  $\beta$ -globin gene cluster appears to have evolved independently in several different tetrapod lineages. In addition to the wealth of data regarding the developmental regulation of hemoglobin synthesis in humans, mice, and other mammals (reviewed by Hardison 2001; Nagel and Steinberg 2001; Brittain 2002), developmental patterns of hemoglobin synthesis have also been characterized in *Xenopus* (Banville and Williams 1985a, 1985b; Fuchs et al. 2006), chicken (Nakazawa et al. 2006; Alev et al. 2008; McIntyre et al. 2008; Nagai and Sheng 2008), and zebra finch (Alev et al. 2009). These data on developmental expression patterns take on special significance in light of our discovery regarding the phylogenetic history of the  $\beta$ -globin gene clusters in each of these taxa. Because the early-expressed  $\beta$ -like globin genes in amphibians, birds, and mammals are not orthologous to one another, our results indicate that stage-specific expression in primitive erythroid cells must have evolved independently in each of these lineages. Likewise, well-characterized functional differences between the products of early- and late-expressed  $\beta$ -like globin genes also must have evolved independently. Although the two  $\beta$ -like globin genes of anole lizard appear to be products of a lizard-specific duplication event (figs. 4 and 6), no data are yet available regarding the developmental expression profiles of these genes. To the best of our knowledge, there are no published reports on developmental patterns of hemoglobin synthesis in any nonavian reptiles. Further study is required to determine whether the two  $\beta$ -like globin genes of lizards have evolved developmentally regulated expression differences.

### Supplementary Material

Supplementary figure S1 is available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

## Acknowledgments

We thank M. Stoeckelhuber for sharing the  $\beta 1$ - and  $\beta 2$ -globin sequences of Indian python, the Broad Institute Genome Sequencing Platform and Genome Sequencing and Analysis Program for permission to use unpublished genome sequence data for *Anolis carolinensis*, and F. Di Palma, K. Lindblad-Toh, and W. Warren from The Genome Center at Washington University for permission to use unpublished genome sequence data for *Taeniopygia guttata*. We are also grateful to S. V. Edwards and two anonymous reviewers for helpful comments on the manuscript. This work was funded by grants to F.G.H. from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq 476739/2008-0), to J.F.S. from the National Institutes of Health, the National Science Foundation, and the Nebraska Research Council, fellowships to T.A.G. by the Studienstiftung des Deutschen Volkes and the Max-Planck-Gesellschaft and partnership in a collaborative grant (“Euroxy”) of the EU’s 6th Framework Programme, and to J.C.O. by the Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT 11080181), Programa Bicentenario en Ciencia y Tecnología (PSD89), and the American Society of Mammalogists.

## References

- Aguileta G, Bielawski JP, Yang Z. 2004. Gene conversion and functional divergence in the  $\beta$ -globin gene family. *J Mol Evol*. 59:177–189.
- Aguileta G, Bielawski JP, Yang Z. 2006. Evolutionary rate variation among vertebrate  $\beta$ -globin genes: implications for dating gene family duplication events. *Gene* 380:21–29.
- Alev C, McIntyre BAS, Nagai H, Shin M, Shinmyozu K, Jakt LM, Sheng G. 2008. BetaA, the major  $\beta$ -globin in definitive red blood cells, is present from the onset of primitive erythropoiesis in chicken. *Dev Dyn*. 237:1193–1197.
- Alev C, Shinmyozu K, McIntyre BAS, Sheng G. 2009. Genomic organization of zebra finch  $\alpha$ - and  $\beta$ -globin genes and their expression in primitive and definitive blood in comparison with globins in chicken. *Dev Genes Evol*. 219:353–360.
- Banville D, Williams JG. 1985a. Developmental changes in the pattern of larval  $\beta$ -globin gene expression in *Xenopus laevis*. Identification of two early larval beta-globin mRNA sequences. *J Mol Biol*. 184:611–620.
- Banville D, Williams JG. 1985b. The pattern of expression of the *Xenopus laevis* tadpole  $\alpha$ -globin genes and the amino acid sequence of the three major tadpole  $\alpha$ -globin polypeptides. *Nucleic Acids Res*. 13:5407–5421.
- Borgese TA, Bertles JF. 1965. Hemoglobin heterogeneity: embryonic hemoglobin in the duckling and its disappearance in the adult. *Science* 148:509–511.
- Brittain T. 2002. Molecular aspects of embryonic hemoglobin function. *Mol Aspects Med*. 23:293–342.
- Brown JL, Ingram VM. 1974. Structural studies on chick embryonic hemoglobins. *J Biol Chem*. 249:3960–3972.
- Bruns GA, Ingram VM. 1973a. Erythropoiesis in the developing chick embryo. *Dev Biol*. 30:455–459.
- Bruns GA, Ingram VM. 1973b. The erythroid cells and haemoglobins of the chick embryo. *Philos Trans R Soc Lond B Biol Sci*. 266:225–305.
- Bulger M, van Doorninck JH, Saitoh N, Telling A, Farrell C, Bender MA, Felsenfeld G, Axel R, Groudine M. 1999. Conservation of sequence and structure flanking the mouse and human  $\beta$ -globin loci: the  $\beta$ -globin genes are embedded within an array of odorant receptor genes. *Proc Natl Acad Sci USA*. 96:5129–5134.
- Burge C, Karlin S. 1997. Prediction of complete gene structures in human genomic DNA. *J Mol Biol*. 268:78–94.
- Burmester T, Haberkamp M, Mitz S, Roesner A, Schmidt M, Ebner B, Gerlach F, Fuchs C, Hankeln T. 2004. Neuroglobin and cytoglobin: genes, proteins, and evolution. *IUBMB Life*. 56:703–707.
- Chan FY, Robinson J, Brownlie A, Shivdasani RA, Donovan A, Brugnara C, Kim J, Lau BC, Witkowska HE, Zon LI. 1997. Characterization of adult  $\alpha$ - and  $\beta$ -globin genes in the zebrafish. *Blood* 89:688–700.
- Cheng JF, Raid L, Hardison RC. 1987. Block duplications of a zeta-zeta-alpha-theta gene set in the rabbit  $\alpha$ -like globin gene cluster. *J Biol Chem*. 262:5414–5421.
- Collins FS, Weissman SM. 1984. The molecular genetics of human hemoglobin. *Prog Nucleic Acid Res Mol Biol*. 31:315–462.
- Cooper SJB, Wheeler D, De Leo A, Cheng J, Holland RAB, Marshall Graves JA, Hope RM. 2006. The mammalian  $\alpha D$ -globin gene lineage and a new model for the molecular evolution of  $\alpha$ -globin gene clusters at the stem of the mammalian radiation. *Mol Phylogenet Evol*. 38:439–448.
- Czelusniak J, Goodman M, Hewett-Emmett D, Weiss ML, Venta PJ, Tashian RE. 1982. Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes. *Nature* 298:297–300.
- De Leo AA, Wheeler D, Lefevre C, Cheng J, Hope R, Kuliwaba J, Nicholas KR, Westerman M, Graves JAM. 2005. Sequencing and mapping hemoglobin gene clusters in the Australian model marsupial *Sminthopsis macroura*. *Cytogenet Genome Res*. 108:333–341.
- Demuth JP, Hahn MW. 2009. The life and death of gene families. *Bioessays* 31:29–39.
- Dodgson JB, Stadt SJ, Choi OR, Dolan M, Fischer HD, Engel JD. 1983. The nucleotide sequence of the embryonic chicken  $\beta$ -type globin genes. *J Biol Chem*. 258:12685–12692.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32:1792–1797.
- Flint J, Tufarelli C, Peden J, et al. (14 co-authors). 2001. Comparative genome analysis delimits a chromosomal domain and identifies key regulatory elements in the  $\alpha$ -globin cluster. *Hum Mol Genet*. 10:371–382.
- Fuchs C, Burmester T, Hankeln T. 2006. The amphibian globin gene repertoire as revealed by the *Xenopus* genome. *Cytogenet Genome Res*. 112:296–306.
- Gillemans N, McMorro T, Tewari R, et al. (12 co-authors). 2003. Functional and comparative analysis of globin loci in pufferfish and humans. *Blood* 101:2842–2849.
- Goh S, Lee YT, Bhanu NV, Cam MC, Desper R, Martin BM, Moharram R, Gherman RB, Miller JL. 2005. A newly discovered human  $\alpha$ -globin gene. *Blood* 106:1466–1472.
- Goodman M, Czelusniak J, Koop BF, Tagle DA, Slightom JL. 1987. Globins: a case study in molecular phylogeny. *Cold Spring Harb Symp Quant Biol*. 52:875–890.
- Goodman M, Moore GW, Matsuda G. 1975. Darwinian evolution in the genealogy of haemoglobin. *Nature* 253:603–608.
- Gorr TA. 1993. Hämoglobin: Sequenz und Phylogenie. Die Primärstruktur von Globinketten des Quastenflossers (*Latimeria chalumnae*) sowie folgender Reptilien: Galapagos -Meerechse (*Amblyrhynchus cristatus*), Grüner Leguan (*Iguana iguana*), Indigonatter (*Dryomarchon corais*), Glatstirnkaiman (*Paleosuchus palpebrosus*). [Ph.D thesis]. [Munich (Germany)]: University of Munich.
- Gorr TA, Mable BK, Kleinschmidt T. 1998. Phylogenetic analysis of reptilian hemoglobins: trees, rates, and divergences. *J Mol Evol*. 47:471–485.

- Hankeln T, Ebner B, Fuchs C, et al. (23 co-authors). 2005. Neuroglobin and cytoglobin in search of their role in the vertebrate globin family. *J Inorg Biochem.* 99:110–119.
- Hankeln T, Wystub S, Laufs T, Schmidt M, Gerlach F, Saaler-Reinhardt S, Reuss S, Burmester T. 2004. The cellular and subcellular localization of neuroglobin and cytoglobin - a clue to their function? *IUBMB Life.* 56:671–679.
- Hardison R. 1998. Hemoglobins from bacteria to man: evolution of different patterns of gene expression. *J Exp Biol.* 201:1099–1117.
- Hardison R. 2001. Organization, evolution, and regulation of the globin genes. In: Steinberg MH, Forget BG, Higgs DR, Nagel RL, editors. Disorders of hemoglobin: genetics, pathophysiology, and clinical management. Cambridge: Cambridge University Press. p. 95–116.
- Hardison R, Miller W. 1993. Use of long sequence alignments to study the evolution and regulation of mammalian globin gene clusters. *Mol Biol Evol.* 10:73–102.
- Hardison RC. 2008. Globin genes on the move. *J Biol.* 7:35.
- Hardison RC, Gelinis RE. 1986. Assignment of orthologous relationships among mammalian  $\alpha$ -globin genes by examining flanking regions reveals a rapid rate of evolution. *Mol Biol Evol.* 3:243–261.
- Hedges SB. 2009. The timetree of life. Oxford: Oxford University Press.
- Hedges SB, Dudley J, Kumar S. 2006. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* 22:2971–2972.
- Hentschel CC, Kay RM, Williams JG. 1979. Analysis of *Xenopus laevis* globins during development and erythroid cell maturation and the construction of recombinant plasmids containing sequences derived from adult globin mRNA. *Dev Biol.* 72:350–363.
- Hiebl I, Braunitzer G, Schneegans D. 1987. The primary structures of the major and minor hemoglobin-components of adult Andean goose (*Chloephaga melanoptera*, Anatidae): the mutation Leu→Ser in position 55 of the  $\beta$ -chains. *Biol Chem Hoppe Seyler.* 368:1559–1569.
- Higgs DR, Vickers MA, Wilkie AO, Pretorius IM, Jarman AP, Weatherall DJ. 1989. A review of the molecular genetics of the human  $\alpha$ -globin gene cluster. *Blood* 73:1081–1104.
- Hoffmann FG, Opazo JC, Storz JF. 2008a. New genes originated via multiple recombinational pathways in the  $\beta$ -globin gene family of rodents. *Mol Biol Evol.* 25:2589–2600.
- Hoffmann FG, Opazo JC, Storz JF. 2008b. Rapid rates of lineage-specific gene duplication and deletion in the  $\alpha$ -globin gene family. *Mol Biol Evol.* 25:591–602.
- Hoffmann FG, Storz JF. 2007. The  $\alpha$ D-globin gene originated via duplication of an embryonic  $\alpha$ -like globin gene in the ancestor of tetrapod vertebrates. *Mol Biol Evol.* 24:1982–1990.
- Hosbach HA, Wyler T, Weber R. 1983. The *Xenopus laevis* globin gene family: chromosomal arrangement and gene structure. *Cell* 32:45–53.
- Hughes JR, Cheng J, Ventress N, Prabhakar S, Clark K, Anguita E, De Gobbi M, de Jong P, Rubin E, Higgs DR. 2005. Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences. *Proc Natl Acad Sci USA.* 102:9830–9835.
- Ikehara T, Eguchi Y, Kayo S, Takei H. 1997. Isolation and sequencing of two  $\alpha$ -globin genes  $\alpha$ (A) and  $\alpha$ (D) in pigeon and evidence for embryo-specific expression of the  $\alpha$ (D)-globin gene. *Biochem Biophys Res Commun.* 234:450–453.
- Jeffreys AJ, Wilson V, Wood D, Simons JP, Kay RM, Williams JG. 1980. Linkage of adult  $\alpha$ - and  $\beta$ -globin genes in *X. laevis* and gene duplication by tetraploidization. *Cell* 21:555–564.
- Jobb G, von Haeseler A, Strimmer K. 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol.* 4:18.
- Kay RM, Harris R, Patient RK, Williams JG. 1980. Molecular cloning of cDNA sequences coding for the major  $\alpha$ - and  $\beta$ -globin polypeptides of adult *Xenopus laevis*. *Nucleic Acids Res.* 8:2691–2707.
- Lynch M. 2007. The origins of genome architecture. Sunderland (MA): Sinauer Associates.
- McIntyre BAS, Alev C, Tarui H, Jakt LM, Sheng G. 2008. Expression profiling of circulating non-red blood cells in embryonic blood. *BMC Dev Biol.* 8:21.
- Nagai H, Sheng G. 2008. Definitive erythropoiesis in chicken yolk sac. *Dev Dyn.* 237:3332–3341.
- Nagel RL, Steinberg MH. 2001. Role of epistatic (modifier) genes in the modulation of the phenotypic diversity of sickle cell anemia. *Pediatr Pathol Mol Med.* 20:123–136.
- Nakatani K, Okuyama H, Shimahara Y, Saeki S, Kim D, Nakajima Y, Seki S, Kawada N, Yoshizato K. 2004. Cytoglobin/STAP, its unique localization in splanchnic fibroblast-like cells and function in organ fibrogenesis. *Lab Invest.* 84:91–101.
- Nakazawa F, Nagai H, Shin M, Sheng G. 2006. Negative regulation of primitive hematopoiesis by the FGF signaling pathway. *Blood* 108:3335–3343.
- Ohno S. 1970. Evolution by gene duplication. Berlin (Germany): Springer-Verlag.
- Opazo JC, Hoffmann FG, Storz JF. 2008a. Differential loss of embryonic globin genes during the radiation of placental mammals. *Proc Natl Acad Sci USA.* 105:12950–12955.
- Opazo JC, Hoffmann FG, Storz JF. 2008b. Genomic evidence for independent origins of  $\beta$ -like globin genes in monotremes and therian mammals. *Proc Natl Acad Sci USA.* 105:1590–1595.
- Opazo JC, Sloan AM, Campbell KL, Storz JF. 2009. Origin and ascendancy of a chimeric fusion gene: the  $\beta/\delta$ -globin gene of paenungulate mammals. *Mol Biol Evol.* 26:1469–1478.
- Patel VS, Cooper SJB, Deakin JE, Fulton B, Graves T, Warren WC, Wilson RK, Graves JAM. 2008. Platypus globin genes and flanking loci suggest a new insertional model for  $\beta$ -globin evolution in birds and mammals. *BMC Biol.* 6:34.
- Pisano E, Cocca E, Mazzei F, Ghigliotti L, di Prisco G, Detrich HW 3rd, Ozouf-Costaz C. 2003. Mapping of  $\alpha$ - and  $\beta$ -globin genes on Antarctic fish chromosomes by fluorescence in-situ hybridization. *Chromosome Res.* 11:633–640.
- Proudfoot NJ, Gil A, Maniatis T. 1982. The structure of the human  $\zeta$ -globin gene and a closely linked, nearly identical pseudogene. *Cell* 31:553–563.
- Reitman M, Grasso JA, Blumenthal R, Lewit P. 1993. Primary sequence, evolution, and repetitive elements of the *Gallus gallus* (chicken)  $\beta$ -globin cluster. *Genomics* 18:616–626.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
- Rücknagel KP, Braunitzer G, Wiesner H. 1988. Hemoglobins of reptiles. The primary structures of the  $\alpha$  I- and  $\beta$  I-chains of common iguana (*Iguana iguana*) hemoglobin. *Biol Chem Hoppe Seyler.* 369:1143–1150.
- Runck AM, Moriyama H, Storz JF. 2009. Evolution of duplicated  $\beta$ -globin genes and the structural basis of hemoglobin isoform differentiation in *Mus*. *Mol Biol Evol.* 26:2521–2532.
- Schmidt M, Gerlach F, Avivi A, et al. (11 co-authors). 2004. Cytoglobin is a respiratory protein in connective tissue and neurons, which is up-regulated by hypoxia. *J Biol Chem.* 279:8063–8069.
- Schmidt M, Laufs T, Reuss S, Hankeln T, Burmester T. 2005. Divergent distribution of cytoglobin and neuroglobin in the murine eye. *Neurosci Lett.* 374:207–211.
- Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W. 2000. PipMaker—a web server for aligning two genomic DNA sequences. *Genome Res.* 10:577–586.
- Stephan S, Adkins R, Anderson J. 2004. Phylogeny and divergence-date estimates of rapid radiations in murid rodents based on multiple nuclear genes. *Syst Biol.* 53:533–553.
- Stoesselhuber M. 1992. Sequenzermittlung von Schlangenhaemoglobinen. Primaerstruktur von Globinketten der Indigonatter (*Drymarchon corais erebennus*) und des Tigerpython (*Python*

- molurus bivittatus*). [diploma thesis]. [Munich (Germany)]: University of Munich.
- Stoekelhuber M, Gorr T, Kleinschmidt T. 2002. The primary structure of three hemoglobin chains from the indigo snake (*Drymarchon corais erebennus*, Serpentes): first evidence for  $\alpha$ D chains and two  $\beta$  chain types in snakes. *Biol Chem*. 383:1907–1916.
- Storz JF, Baze M, Waite JL, Hoffmann FG, Opazo JC, Hayes JP. 2007. Complex signatures of selection and gene conversion in the duplicated globin genes of house mice. *Genetics* 177:481–500.
- Storz JF, Hoffmann FG, Opazo JC, Moriyama H. 2008. Adaptive functional divergence among triplicated  $\alpha$ -globin genes in rodents. *Genetics* 178:1623–1638.
- Tatusova TA, Madden TL. 1999. BLAST 2 sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett*. 174:247–250.
- Taylor JS, Raes J. 2004. Duplication and divergence: the evolution of new genes and old ideas. *Annu Rev Genet*. 38:615–643.
- Townes TM, Shapiro SG, Wernke SM, Lingrel JB. 1984. Duplication of a four-gene set during the evolution of the goat  $\beta$ -globin locus produced genes now expressed differentially in development. *J Biol Chem*. 259:1896–1900.
- Wheeler D, Hope R, Cooper SB, Dolman G, Webb GC, Bottema CD, Gooley AA, Goodman M, Holland RA. 2001. An orphaned mammalian  $\beta$ -globin gene of ancient evolutionary origin. *Proc Natl Acad Sci USA*. 98:1101–1106.
- Wheeler D, Hope RM, Cooper SJB, Gooley AA, Holland RAB. 2004. Linkage of the  $\beta$ -like  $\omega$ -globin gene to  $\alpha$ -like globin genes in an Australian marsupial supports the chromosome duplication model for separation of globin gene clusters. *J Mol Evol*. 58:642–652.
- Zhang J. 2003. Evolution by gene duplication: An update. *Trends Ecol Evol*. 18:292–298.