June 2007

The α^D-Globin Gene Originated via Duplication of an Embryonic α-Like Globin Gene in the Ancestor of Tetrapod Vertebrates

Federico G. Hoffmann
University of Nebraska - Lincoln, fhoffmann2@unl.edu

Jay F. Storz
University of Nebraska - Lincoln, jstorz2@unl.edu

Published in *Molecular Biology and Evolution* **24**:9 (2007), pp. 1982-1990; doi 10.1093/molbev/msm127

Accepted for publication June 18, 2007; published online June 22, 2007. Associate editor for *MBE*: William Jeffery.

The α^D-Globin Gene Originated via Duplication of an Embryonic α-Like Globin Gene in the Ancestor of Tetrapod Vertebrates

Federico G. Hoffmann and Jay F. Storz

School of Biological Sciences, University of Nebraska–Lincoln (*Correspondence email:* jstorz2@unl.edu)

Abstract

Gene duplication is thought to play an important role in the co-option of existing protein functions to new physiological pathways. The globin superfamily of genes provides an excellent example of the kind of physiological versatility that can be attained through the functional and regulatory divergence of duplicated genes that encode different subunit polypeptides of the tetrameric hemoglobin protein. In contrast to prevailing views about the evolutionary history of the α-globin gene family, here we present phylogenetic evidence that the α^A- and α^D-globin genes are not the product of a single, tandem duplication of an ancestral globin gene with adult function in the common ancestor of extant birds, reptiles, and mammals. Instead, our analysis reveals that the α^D-globin gene of amniote vertebrates arose via duplication of an embryonic α-like globin gene that predated the radiation of tetrapods. The important evolutionary implication is that the distinct biochemical properties of α^D-hemoglobin (*HbD*) are not exclusively derived characters that can be attributed to a post-duplication process of neofunctionalization. Rather, many of the distinct biochemical properties of *HbD* are retained ancestral characters that reflect the fact that the α^D-globin gene arose via duplication of a gene that had a larval/embryonic function. These insights into the evolutionary origin of *HbD* illustrate how adaptive modifications of physiological pathways may result from the retention and opportunistic co-option of ancestral protein functions.

Keywords: hemoglobin, gene duplication, alpha-D globin, neofunctionalization, molecular phylogenetics, gene family evolution

Supplemental Material: Supplemental material appears following the "References"; it is also available at *Molecular Biology and Evolution* online, http://mbe.oxfordjournals.org/

Introduction

Gene duplication can result in the acquisition of novel protein functions in cases where one duplicate copy retains the original function of the ancestral gene while the other copy accumulates new mutations that adapt the encoded protein to a new or modified physiological task (Ohno 1970; Kimura and Ohta 1974; Goodman *et al.* 1975; Ohta 1993, 1994; Hughes 1999). This functional specialization may entail changes in protein function as well as changes in the tissue specificity or developmental timing of gene expression. The globin superfamily of genes provides an excellent example of the kind of physiological versatility that can be attained through the functional and regulatory divergence of duplicated genes that encode different subunit polypeptides of the tetrametic hemoglobin protein (Goodman *et al.* 1975; Hardison 2001). In amniote vertebrates, hemoglobin has been optimized for oxygen transport under the vastly different physiological conditions encountered during the embryonic, fetal, and adult stages of development. For example, in eutherian mammals, regulatory switching between paralogous genes that encode functionally distinct subunit polypeptides can produce adult hemoglobins that are optimized for pulmonary/tissue oxygen transport, fetal hemoglobins that are optimized for placental/tissue oxygen transport, and embryonic hemoglobins that are optimized for diffusional oxygen scavenging from the amniotic fluid as well as oxygen transport during the transition from a vitelline to placental circulation (Hardison 1998, 2001; Nagel and Steinberg 2001; Brittain 2002).

In gnathostome vertebrates, the hemoglobin protein is a heterotetramer composed of two α-chain and two β-chain subunits. In amniote vertebrates, each of the different subunit polypeptides is encoded by different sets of duplicated genes that are located on different chromosomes. In amphibians and ray-finned fishes, the α- and β-like globin genes are closely linked on the same chromosome, reflecting the fact that the progenitors of the α- and β-globin gene families arose via tandem duplication of an ancestral globin gene approximately 450-500 mya, prior to the radiation of amniote vertebrates (Goodman *et al.* 1975; Czelusniak *et al.* 1982; Goodman *et al.* 1987).

Members of the α- and β-globin gene families have diversified in both biochemical properties and developmental timing of expression. The genes in each cluster are generally organized in a 5'-3' sequence that is co-linear with the temporal sequence of expression during development. The archetypal arrangement of the tetrapod α-globin gene family consists of two functional genes, each of which may be present in one or more copies: an α-like globin gene expressed early in embryonic development, α^E-globin (known as α^L-globin in amphibians, ζ-globin in mammals, and π-globin in birds), and the adult α^A-globin gene. Birds, mammals, and reptiles possess an additional member of the family, the adult α^D-globin gene, whereas the θ-globin gene is found only in mammals. In nearly all mammals studied to date, the α-chains of adult hemoglobin are encoded by duplicate copies of the α^A-globin gene, which are almost always identical in sequence and therefore encode identical polypeptides (Zimmer *et al.* 1980; Higgs *et al.* 1989). The α^D-globin gene was only recently identified in mammals (Goh *et al.* 2005; Hughes *et al.* 2005; Cooper *et al.* 2006), and although it is transcriptionally active, the product of this gene is not known to be assembled into functional hemoglobin tetramers. In birds and reptiles, by contrast, the α-chains of adult hemoglobin are encoded by both the α^D- and α^A-globin genes. In most birds studied to date, α^D-containing hemoglobin

(*HbD*) constitutes the minor fraction of adult hemoglobin and α^A-containing hemoglobin (*HbA*) constitutes the major fraction (Borgese and Bertles 1965; Brown and Ingram 1974; Hiebl *et al.* 1987). Thus, unlike the case in most mammals, the mature erythrocytes of adult birds and reptiles may contain a mixture of functionally distinct hemoglobin isoforms that have different biochemical properties. Specifically, *HbD* generally has a higher oxygen affinity and a higher cooperativity of oxygen binding than *HbA* (Cirotto and Geraci 1975; Baumann *et al.* 1984; Riggs 1998; Knapp *et al.* 1999). Thus, regulatory adjustments that alter the stoichiometric ratio of these two isoforms in circulating erythrocytes may modulate rates of oxygen flux in response to changes in metabolic demand (Hiebl *et al.* 1987; Hiebl *et al.* 1988).

It has traditionally been thought that the α^D- and α^A-globin genes arose via tandem duplication of an ancestral proto α^A-globin gene with adult function in the common ancestor of birds and mammals (Czelusniak *et al.* 1982; Hardison 1998, 2001), and recent phylogenetic analyses lend support to this idea (Cooper *et al.* 2006). According to this scenario, biochemical properties of α^D-globin that distinguish it from α^A-globin may be derived characteristics that evolved under the influence of directional selection that favored a physiological division of labor between the two co-expressed gene duplicates. This scenario has intriguing evolutionary implications because functional studies have demonstrated striking similarities between *HbD* and embryonic hemoglobin (*HbE*) with respect to oxygen-binding affinity, cooperativity of oxygen binding, solubility, and responsiveness to chloride ions and other allosteric effectors (Cirotto and Geraci 1975; Baumann *et al.* 1984; Chapman *et al.* 1980; Chapman *et al.* 1982; Grigg *et al.* 1993; Knapp *et al.* 1999). If the α^D- and α^A-globins arose via tandem duplication of an ancestral proto α^A-globin gene with adult function, then it would appear that α^D-globin has convergently evolved a suite of biochemical properties that are characteristic of embryonic globins.

In contrast to prevailing views about the evolutionary history of the α-globin gene family, here we present evidence that the α^D- and α^A-globins are not the product of a single, tandem duplication of an ancestral globin gene with adult function in a common ancestor to extant birds, reptiles, and mammals. Instead, results of our phylogenetic analyses reveal that the α^D-globin of amniote vertebrates arose via duplication of a proto α^E-globin gene that predated the radiation of tetrapods. The important evolutionary implication is that the distinct biochemical properties of *HbD* are not exclusively derived characters that evolved during a post-duplication process of neofunctionalization. Rather, many of the distinct biochemical properties of *HbD* are retained ancestral characters that reflect the fact that the α^D-globin gene arose via duplication of a gene that had a larval/embryonic function in the ancestor of all tetrapods.

Material and Methods

DNA Sequence Data

We obtained DNA sequences for structural genes in the α-globin gene family from either Genbank or Ensembl (release 37, February 2006). When possible, we focused on sequences from a single genomic contig, genomic scaffold, or full chromosome, depending on the nature of the available data. We also included sequences from shorter records based on genomic DNA or cDNA in order to attain a broad and balanced taxonomic coverage of the tetrapod phylogeny. This provision allowed us to include sequences from amphibians (*Xenopus* and *Pleurodeles*), reptiles (*Geochelone* and *Hydrophis*), birds (*Cairina*), as well as some additional mammalian taxa. The basic annotation was derived from the database records in most cases, but we also identified globin genes in unannotated sequences by comparing known exon sequences to genomic contigs using the program BLAST 2 sequences (Version 2.2; Tatusova and Madden 1999), available from the NCBI website: http://www.ncbi.nlm.nih.gov/blast/bl2seq .

We restricted the phylogenetic analysis to functional copies of the four α-like globin genes and we excluded redundant sequences. We obtained sequences from functional members of the α-globin gene family in representatives from all tetrapod classes including amphibians, reptiles, birds, and mammals. The α-globin sequences from zebra fish (*Danio rerio*) and puffer fish (*Sphoeroides nephelus*) were used as outgroups. Alignment was based on the amino acid translation, and was carried out in ClustalW (Thompson *et al.* 1994) as implemented in BioEdit (Hall 1999). Having confirmed orthologous relationships among the larval/embryonic α-like globins among all vertebrate taxa (see Results), we use the name α^E-globin to refer to the α^L-globin of amphibians, the π-globin of birds, and the ζ-globin of mammals.

Amino Acid Sequence Data

Due to the limited amount of DNA sequence data available for the α-like globin genes of reptiles, we also assembled a protein sequence data set to corroborate observations derived from the DNA data set. In the protein data set, we increased our representation of α-like globin genes from amphibians and reptiles, and we included translated amino acid sequences from a subset of the avian and mammalian taxa that were included in the DNA analyses. Specifically, we included translated sequences of α^E-, α^D-, and α^A-globin from two representatives of the class Aves (duck and chicken) and translated sequences of α^E-, α^D-, α^A-, and θ-globin from four representatives of the class Mammalia (an Australian marsupial, a New world marsupial, cow, and human). As with the DNA analysis, we included translated α-globin sequences from zebra fish and puffer fish as outgroups.

Phylogenetic Inference

We estimated phylogenetic relationships among the different α-like globin DNA sequences in our data set using neighbor-joining (NJ), maximum parsimony (MP), maximum likelihood (ML), and Bayesian approaches. Neighbor-joining and MP searches were conducted in PAUP version 4.b10 (Swofford 2002), ML analyses were implemented in Treefinder version May 2006 (Jobb *et al.* 2004), and Bayesian analyses were implemented in MrBayes version 3.1.2 (Ronquist and Huelsenbeck 2003). For the the NJ analyses, we selected the best-fit model of nucleotide substitution by using hierarchical Likelihood Ratio Tests as implemented in the program ModelTest version 3.7 (Posada and Crandall 1998). For the the MP analyses, all substitutions were equally weighted. In ML and Bayesian analyses, we allowed searches to jointly optimize branch lengths and parameter values for the best-fit model of nucleotide substitution. We then compared tree searches with and

Table 1 Results from the Parametric Bootstrap Test. We Simulated 1000 Sequence Data Sets Using the Tree and GTR + Γ Model of Nucleotide Substitution that Corresponded to the Null Hypothesis. The Difference in Likelihood Score Between the Null Hypothesis ML Topology and the Alternative Hypothesis ML Topology is given as $\Delta = (-\ln L\ H_0) - (-\ln L\ H_{alt})$

Topologies Compared	Mean Δ from Simulated Samples	Observed Δ	Critical Δ values (α = 0.01)	P-value
H_0 = Fig 2B H_{alt} = Fig 2A	1.79	−2.79	−1.50	$P < 0.003$
H_0 = Fig 2C H_{alt} = Fig 2A	22.11	−11.33	5.13	$P < 0.001$
H_0 = Fig 2D H_{alt} = Fig 2A	38.65	−12.93	13.24	$P < 0.001$

without site-specific rates for each of the three codon positions using an independent General-Time-Reversible model (GTR; Rodriguez *et al.* 1990) in which rate variation followed a discrete gamma distribution for each position. Support for nodes in NJ, MP, and ML analyses was evaluated using the bootstrap procedure. We employed 1000 pseudoreplicates in NJ and MP analyses, and 100 in ML analyses using the same parameter settings that were used in the phylogenetic searches. Bayesian analyses were run for 5×10^6 iterations of a Markov Chain Monte Carlo algorithm, with samples taken every 1000 iterations. Support for the nodes and parameter estimates were derived from a majority rule consensus of the last 2,500 trees sampled after convergence. In the case of the protein sequence data, we used ProtTest version 1.3 (Abascal *et al.* 2005) to find the best-fitting model of amino acid substitution, and set ML searches in Treefinder to jointly estimate the phylogeny and parameter values for the selected model.

Hypothesis Testing

To compare alternative tree topologies, we used the Shimodaira-Hasegawa topology test (Shimodaira and Hasegawa 1999), and the approximately unbiased test suggested by Shimodaira (2002). We computed site-specific likelihood scores for each site in the baseml program (Yang 1997), and used the CONSEL program (Shimodaira and Hasegawa 2001) to test whether differences among alternative topologies were statistically significant. Given that the hypotheses evaluated were mutually exclusive, we performed a SOWH test (Swofford *et al.* 1996), as suggested by Goldman *et al.* (2000). This test uses a parametric bootstrapping approach to test for differences in topology. First we constrained our searches to find the ML tree that was congruent with the null hypothesis and its associated model of nucleotide substitution, partitioning data into the three different codon positions. Next, we simulated 1,000 data sets based on the null hypothesis ML phylogram, and its corresponding model of nucleotide substitution. For each of the 1,000 simulated data sets we calculated the difference in likelihood score, Δ, between the null hypothesis ML topology and the alternative hypothesis ML topology. Using an α-level of 0.01, the null hypothesis ML topology is rejected if ≥99% of the simulation-based Δ-values exceed the observed value.

Results

Sequence Data

We collected 120 functional α-like globin sequences from a diverse array of tetrapod taxa, including amphibians, reptiles, birds, and mammals, in addition to two fish (zebra fish HBAA1 and puffer fish HBA2). The final data set consisted of a 429 bp alignment excluding stop codons (Supplementary Material, Table 1).

Phylogenetic Inference

On the basis of previous work, the following phylogenetic results are predicted: (*i*) α^E-globins from mammals and birds, α^D-globin, and θ-globin should each form a monophyletic clade; (*ii*) α^A-globin should be paraphyletic relative to θ-globin; (*iii*) mammalian α^A-globin should be sister to mammalian θ-globin; and (*iv*) relationships within each set of orthologous sequences should match the expected phylogenetic relationships among species. The goals of our phylogenetic analysis were to resolve relationships among the four α-like globin paralogs and to assess the robustness of the inferred tree in comparison with trees predicted by alternative hypotheses regarding the evolution of the α-globin gene family. Neighbor-joining, MP, ML, and Bayesian phylogenetic analyses of our DNA alignment produced highly concordant results. Here, we present results from the ML analysis, with support from ML bootstrap (MLbs) and Bayesian posterior probabilities (BApp). Results of the NJ and MP analyses are available in the supplementary material (Supplementary Material, Figs 1 and 2). In all cases the sequences were grouped into three main lineages: fish α-globins (which were sister to tetrapod α-globins), tetrapod α^E- and α^D-globins, and tetrapod α^A- and θ-globins (Figure 1). The ML analysis of the protein data set was consistent with results based on the DNA sequences (Supplementary Material, Figure 3).

We found strong support for monophyly of the clade containing all α^E-globin sequences (MLbs = 78%, BApp = 98%). Surprisingly, we also found support for a sister-group relationship between a clade that included all embryonic α^E-globin sequences (amphibian α^L-globin, avian π-globin, and mammalian ζ-globin), and a clade that included all α^D-globin sequences in addition to an α^A-globin gene of the slender sea-snake (MLbs = 59%, and BApp = 99%). Bootstrap support for the monophyly of the clade that grouped all α^D-globin genes with the slender sea-snake α^A-globin was over 70% in ML, and the posterior probability was just below the 95% threshold set in the Bayesian analysis. Although the putative α^D-globin gene of the slender sea-snake grouped with other α^D-globin sequences in the phylogeny reconstruction, this sequence does not share a second-codon deletion that is shared by all other α^D-globin genes. Instead, the second codon of the sea-snake sequence codes for valine, as is the case for all amniote α^A-globin genes in our database. This suggests a history of gene conversion where the 5' end of the sea-snake α^D-globin coding sequence has been converted by the α^A-globin paralog. Further insights into the na-

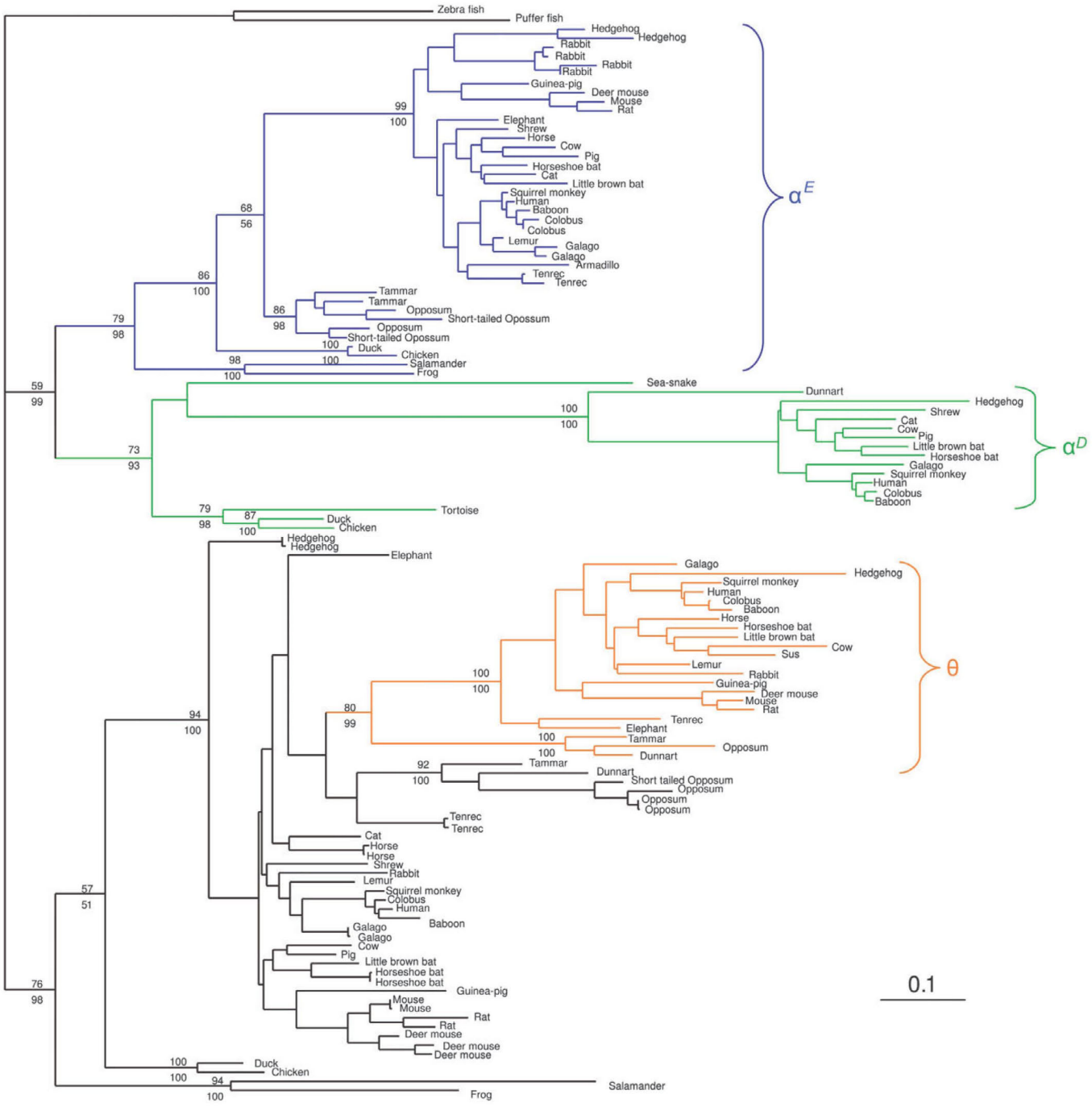

Figure 1. Maximum likelihood phylogram describing relationships among the α-like globin genes of vertebrates, with α^E-globin sequences (including amphibian α^L-globin, avian π-globin, and mammalian ζ-globin) shown in blue, α^D-globin sequences shown in green, α^A-globin sequences shown in black, and θ-globin sequences shown in red. Two fish α-globin sequences were used as outgroups. Searches were conducted in Treefinder version May 2006 (Jobb, von Haeseler, and Strimmer 2004) using an independent site-specific GTR+Γ model of nucleotide substitution for each codon position. Numbers above the nodes correspond to maximum likelihood bootstrap support values, and those below the nodes correspond to Bayesian posterior probabilities.

ture and prevalence of interparalog gene conversion between the α^A- and α^D-globin genes should be possible when genomic sequence data become available for snakes and other squamate reptiles.

We also found strong support for a clade that included amphibian, reptilian, and avian α^A-globins, as well as the mammalian α^A- and θ-globins (MLbs = 76%, BApp = 98%). As expected, the α^A-globin clade was found to be paraphyletic rel-

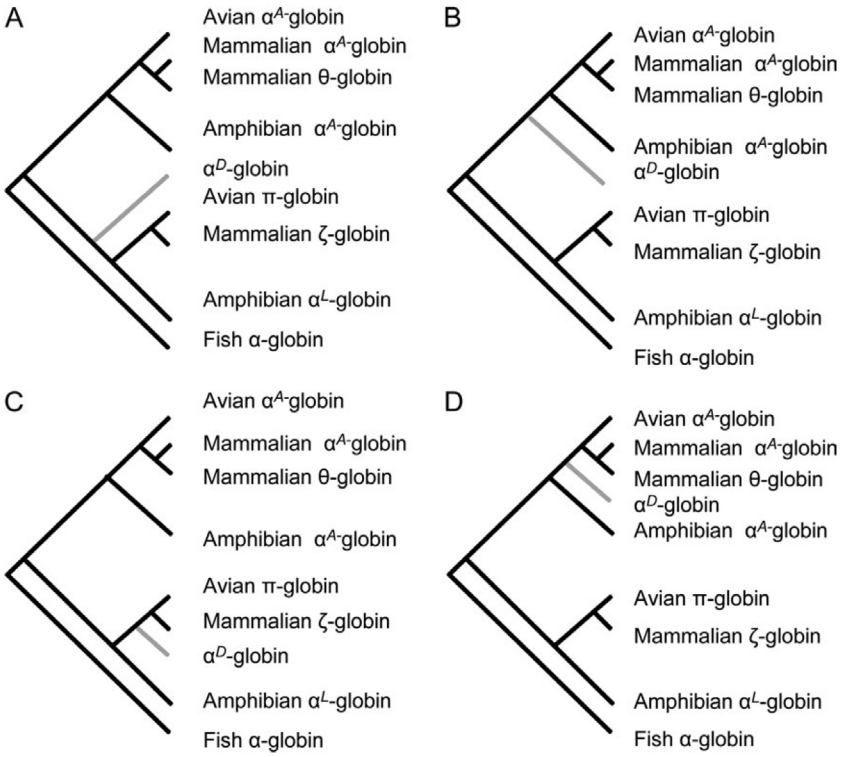

Figure 2. Schematic representations of alternative hypotheses regarding phylogenetic relationships among the α-like globin genes of amniote vertebrates. In A, α^D-globin arose via duplication of an ancestral embryonic α^E-globin that predated the radiation of tetrapods. In B, the α^D-globin arose via duplication of an ancestral adult α^A-globin that predated the radiation of tetrapods. Our results support hypothesis A. In C, α^D-globin arose via duplication of an ancestral embryonic α^E-globin that occurred after the divergence of amphibians from the common ancestor of amniote vertebrates, whereas in D, α^D-globin arose via duplication of an ancestral adult α^A-globin that occurred after the divergence of amphibians from the common ancestor of amniote vertebrates.

ative to mammalian θ-globin. The monophyly of the mammalian θ-globins, in turn, was strongly supported (MLbs = 80%, BApp = 99%).

Hypotheses Testing

The topology of the ML tree in Figure 1 met a subset of our initial expectations regarding phylogenetic relationships within and among the different paralogs, although there were several minor discrepancies between the observed and expected relationships among mammalian taxa within each set of orthologous sequences. Eutherian and marsupial sequences were placed sister to each other in the case of α^E-globin, α^D-globin, and θ-globin. Adult α^A-globins, however, departed from this pattern. Here, the deepest nodes corresponded to the split of the hedgehog α^A-globins from a clade that contained the remaining therian α^A-globins and therian θ-globins. Marsupial α^A-globins were sister to the two tenrec paralogs, and this clade in turn was sister to the θ-globin clade. Statistically significant support for the sister relationship between therian θ-globins and the clade containing marsupial and tenrec α^A-globins was only found in Bayesian analyses (BApp = 99%, MLbs < 50%). The presence of θ-globin sequences in both marsupial and eutherian mammals indicates that a copy of this gene was present in the therian lineage prior to the divergence of marsupials from eutherians. The lack of support for a sister relationship between the therian α^A- and θ-globin sequences is most likely attributable to lack of power and/or long branch attraction between marsupial α^A-globins and therian θ-globins.

The main discrepancy between our inferred tree and the tree predicted by the conventional view of α-globin gene family evolution concerns the placement of the α^D-globin lineage. Our inferred tree indicates that α^D-globin is more closely related to embryonic α^E-globin than to adult α^A-globin, and that the duplication that gave rise α^D-globin preceded the radiation of extant tetrapods.

Given that our results contradict prevailing views of α-globin gene family evolution, we used constrained searches to explore differences in likelihood scores among the alternative hypotheses (Figure 2). Specifically, we explored alternative placements of the α^D-globin sequences that are predicted under competing evolutionary hypotheses, and compared differences among the alternative tree topologies in a statistical framework. The conventional view is that the duplication that gave rise to α^D-globin occurred after the divergence of amphibians from the common ancestor of all amniote vertebrates (Figure 3A). By contrast, our results suggest that this duplication took place in the common ancestor of all extant tetrapods (Figure 3B). Accordingly, we carried out ML phylogenetic reconstructions in which α^D-globin divergence occurred before the split of either the α^E- or α^A globins of amphibians (Figure 2A and 2B), and compared these inferred trees to constrained

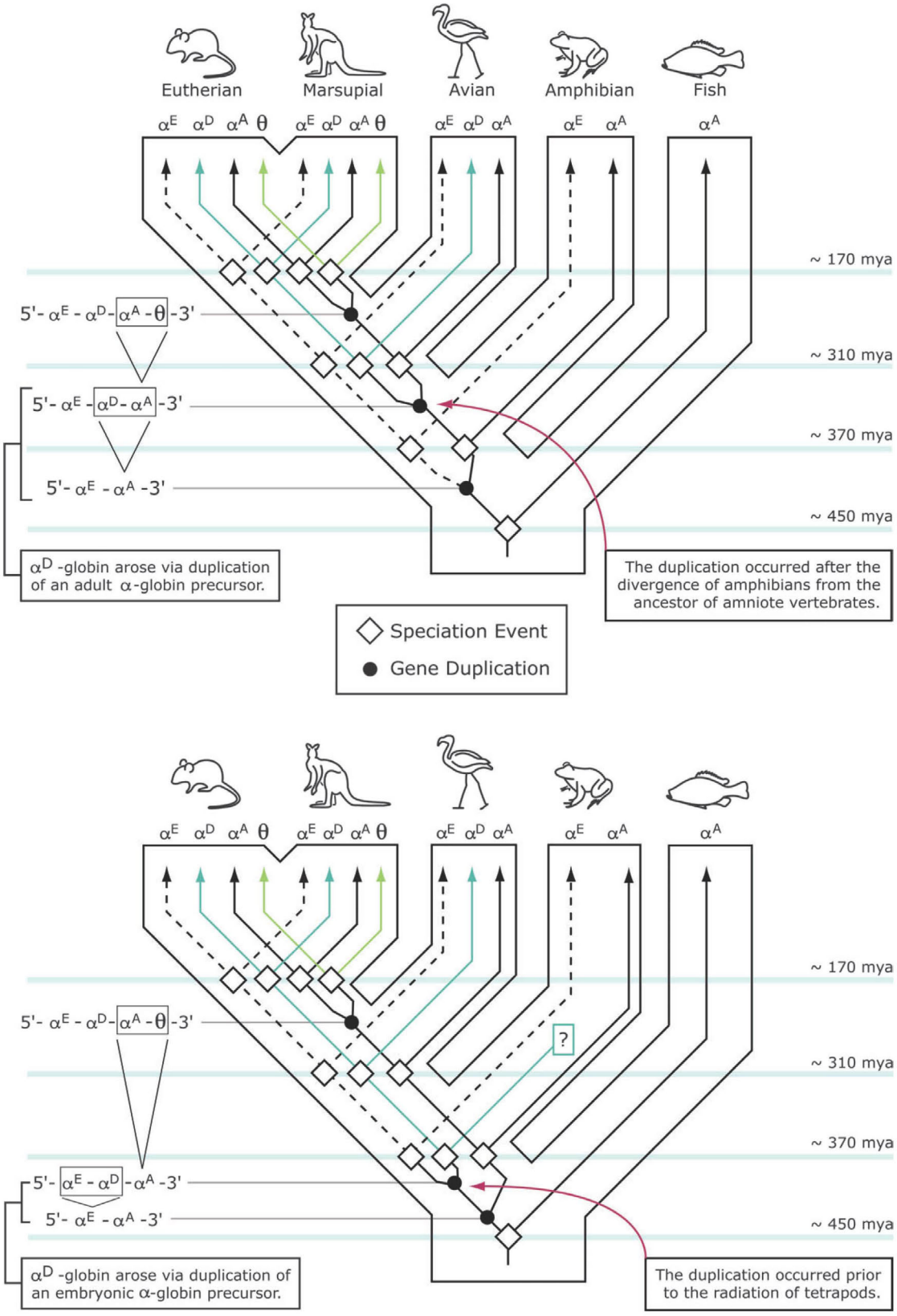

Figure 3. Two alternative hypotheses regarding the evolutionary history of the α-globin gene family. Panel A summarizes the prevailing view, in which the α^E-globin and α^A-globin genes arose via tandem duplication of the α-globin pro-ortholog in the ancestor of all extant tetrapods. This was then followed by a second duplication of the adult α^A-globin gene which gave rise to the α^D-globin gene in the common ancestor of all amniotes. Panel B shows the evolutionary scenario favored by our phylogenetic analyses. Under this scenario, the α^D-globin gene arose via duplication of the α^E-globin pro-ortholog in the common ancestor to all tetrapods.

trees in which amphibian α^E- or α^A globins arose prior to the divergence of α^D-globin (Figure 2C and 2D). Differences in likelihood scores associated with each of the four alternative hypotheses were not statistically significant in the SH and AU tests. However, the parametric-bootstrapping SOWH test was able to distinguish the best ML tree (Figure 2A) from each of the three alternative constrained trees with a high level of statistical significance ($P < 0.005$; Table 1).

Discussion

We used a large number of vertebrate genome sequences and recently developed phylogenetic methods to reevaluate evolutionary relationships among members of the α-globin gene family in vertebrates. Results of our phylogenetic analyses indicate that α^D-globin is more closely related to embryonic α^E-globin than to adult α^A-globin, and that the tandem duplication that gave rise to α^D-globin preceded the radiation of extant tetrapods. Our conclusion that the α^D-globin gene arose via duplication of a proto α^E-globin gene has important evolutionary implications for understanding the role of hemoglobin isoform differentiation in the oxygen-transport systems of birds and other archosaurs.

The prevailing view of α-globin gene family evolution has been that the α^E- and α^A-globin genes arose via tandem duplication of the α-globin pro-ortholog in the ancestor of all extant tetrapods (Czelusniak *et al.* 1982; Hardison 1991, 2001; Aguileta *et al.* 2006; Cooper *et al.* 2006; Figure 2A). This was then followed by a second duplication of the adult α^A-globin gene, which gave rise to α^D-globin. This second duplication was initially thought to be restricted to modern reptiles and birds, but the recent discovery of a transcriptionally active copy of α^D-globin in the α-globin gene cluster of mammals (Goh *et al.* 2005; Hughes *et al.* 2005; Cooper *et al.* 2006) suggests that the duplication must have occurred in the common ancestor of amniote vertebrates. Several previous studies have presented phylogenetic reconstructions consistent with a sister relationship between the α^D- and α^E-globins (Goodman *et al.* 1982; Fushitani *et al.* 1996; Gorr *et al.* 1998; Wheeler *et al.* 2004), but the complete history of gene duplication and divergence was never reconstructed nor were the evolutionary implications explored.

If the duplication leading to α^D-globin had occurred after the lineage leading to amniotes diverged from the stem lineage of amphibians (the conventional view), we would expect α^D-globin sequences to be basal to either the amniote α^E- or amniote α^A-globin clade, and to branch after the divergence of the amphibian α^E- or α^A-globins (Figure 2A). However, monophyly of the group that includes the α^E-globins to the exclusion of α^D-globins is strongly supported in both ML and Bayesian analyses, as is the monophyly of the clade that includes the α^A-globins of all amniotes and the θ-globins of mammals to the exclusion of α^D-globins. We therefore conclude that the gene duplication that gave rise to α^D-globin predated the deepest split within either the α^E- or α^A- globin clades, as illustrated in Fig 3B. This would imply that an α^D-globin paralog was present in the ancestor of all tetrapods, regardless of whether α^D-globin is allied to embryonic α^E- or adult α^A-globins. To date, there is no evidence for the presence of an α^D-globin gene in amphibians, which suggests that it may have been secondarily lost.

Our evolutionary hypothesis that α^D-globin is the product of a duplication of the α^E-globin pro-ortholog is supported by analyses of both DNA and amino acid sequences, which indicate a sister relationship between the α^D- and α^E-globins (Figure 3B). Support for this relationship is strong in Bayesian analyses (BApp = 99%), and moderate in the ML analysis (MLbs = 59%). The sister relationship of α^D- and α^E-globins is also strongly supported by the parametric-bootstrapping SOWH test. This indicates that the α^D-globin gene, which is expressed in adult birds and reptiles, shares a common ancestor with α^E-globin, which is only expressed during the earliest stages of embryonic development. Our analyses also provide strong support for monophyly of the clade that includes all α^E-globins (amphibian α^L-globins, mammalian ζ-globins, and avian π-globins) to the exclusion of all other α-like globin sequences. This suggests an orthologous relationship among all α-like globin genes that are expressed during the earliest stages embryonic development, and is consistent with previous studies that confirmed an orthologous relationship between avian π-globins and mammalian ζ-globins (Czelusniak *et al.* 1982; Proudfoot *et al.* 1982).

In the common ancestor of amniote vertebrates, it may be that the ancestral α^D-globin was expressed early in development, and was later recruited for expression in the mature erythrocytes of adult birds and reptiles, although the ancestral mode of embryonic expression has been retained in some avian species (Godovac-Zimmermann and Braunitzer 1984). In most birds, the α^D-globin gene is expressed at lower levels than α^A-globin, but the relative expression levels of the two genes vary widely among different species (Borgese and Bertles 1965; Godovac-Zimmermann and Braunitzer 1984, 1985; Hiebl *et al.* 1986; Hiebl *et al.* 1987; Hiebl *et al.* 1989; Nothum, Braunitzer *et al.* 1989; Nothum, Weber *et al.* 1989; Ikehara *et al.* 1997).

The functional differentiation of *HbA* and *HbD* isoforms appears to have played an important role in the evolution of hypoxia tolerance in a number of birds, including high-soaring vultures (Hiebl *et al.* 1988; Weber *et al.* 1988; Hiebl *et al.* 1989) and migratory ducks and geese (Hiebl *et al.* 1986; Hiebl *et al.* 1987). The graded oxygen affinities of α^A- and α^D- containing hemoglobin isoforms provides a 'cascade' mechanism for fine-tuning hemoglobin-oxygen affinity in response to variation in ambient oxygen tension. Under conditions of high-altitude hypoxia, high-affinity *HbD* isoforms ensure efficient pulmonary oxygen loading in the parabranchial lung whereas low-affinity *HbA* isoforms ensure the efficient unloading of oxygen to the cells of aerobically metabolizing tissues (Hiebl *et al.* 1988; Weber *et al.* 1988).

One of the most striking examples of the role of hemoglobin isoform differentiation in high-altitude respiration involves Rüppell's griffon (*Gyps rueppelli*), an African vulture that nests at sea-level and is known to fly at altitudes of over 11,000 m. As a result of tandem duplication and divergence of the α^A- and α^D-globin genes, the red blood cells of these birds contain a heterogeneous mixture of four functionally distinct hemoglobin isoforms with the following rank-order of oxygen-binding affinities (low to high): *HbA*, *HbA'*, *HbD*, and *HbD'*. The graded oxygen affinities of these four hemoglobin iso-

forms permits a cascade mechanism of pulmonary/tissue oxygen transport, and appears to provide an important regulatory reserve of oxygen transport capacity (Hiebl *et al.* 1988; Weber, Hiebl, and Braunitzer 1988). The fact that the α^D-globin gene arose via duplication of a precursor gene with an embryonic/larval function suggests that the encoded protein was 'pre-adapted' to the task of pulmonary oxygen loading at low partial pressures of oxygen. The protein was pre-adapted to this new task in the sense that the requisite biochemical properties were present in the ancestral condition, and could therefore be co-opted for a modified function when the possession of a high-affinity adult hemoglobin became advantageous. Given what we know about the role of *HbD* in the evolution of hypoxia tolerance in birds, and possibly in other archosaurs (Shishikura 2002; Stoeckelhuber, Gorr *et al.* 2002), our insights into the evolutionary origin of the α^D-globin gene illustrate how adaptive modifications of physiological pathways may result from the retention and opportunistic co-option of ancestral protein functions.

Acknowledgments

We thank G. Ortí, E. Gering, A. Runck, J. C. Opazo, E. P. Lessa, W. Jeffery, and two anonymous reviewers for helpful comments and suggestions. This work was funded by a Postdoctoral Fellowship in Population Biology to FGH from the University of Nebraska–Lincoln, an NSF grant to JFS (DEB-0614342), a Layman Award to JFS, and an Interdisciplinary Research Grant to JFS from the Nebraska Research Council.

References

Abascal F, Zardoya R, Posada D. Prottest: selection of best-fit models of protein evolution. Bioinformatics (2005) 21:2104–2105.

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

Baumann R, Goldbach E, Haller EA, Wright PG. Organic phosphates increase the solubility of avian haemoglobin D and embryonic chicken haemoglobin. Biochem J (1984) 217:767–771.

Borgese TA, Bertles JF. Hemoglobin heterogeneity: embryonic hemoglobin in the duckling and its disappearance in the adult. Science (1965) 148:509–511.

Brittain T. Molecular aspects of embryonic hemoglobin function. Mol Aspects Med (2002) 23:293–342.

Brown JL, Ingram VM. Structural studies on chick embryonic hemoglobins. J Biol Chem. (1974) 249:3960–3972.

Chapman BS, Hood LE, Tobin AJ. Minor early embryonic chick hemoglobin M. Amino acid sequences of the epsilon and alpha D chains. J Biol Chem. (1982) 257:651–658.

Chapman BS, Tobin AJ, Hood LE. Complete amino acid sequences of the major early embryonic alpha-like globins of the chicken. J Biol Chem. (1980) 255:9051–9059.

Cirotto C, Geraci G. Embryonic chicken hemoglobins. studies on the oxygen equilibrium of two pure components. Comp Biochem Physiol A (1975) 51:159–163.

Cooper SJB, Wheeler D, De Leo A, Cheng J, Holland RAB, Marshall Graves JA, Hope RM. The mammalian alphaD-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. (2006) 38:439–448.

Czelusniak J, Goodman M, Hewett-Emmett D, Weiss ML, Venta PJ, Tashian RE. Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes. Nature (1982) 298:297–300.

Fushitani K, Higashiyama K, Moriyama EN, Imai K, Hosokawa K. The amino acid sequences of two alpha chains of hemoglobins from Komodo dragon *Varanus komodoensis* and phylogenetic relationships of amniotes. Mol Biol Evol. (1996) 13:1039–1043.

Godovac-Zimmermann J, Braunitzer G. Hemoglobin of the adult white stork (*Ciconia ciconia*, Ciconiiformes). the primary structure of alpha a- and beta-chains from the only present hemoglobin component. Hoppe Seylers Z Physiol Chem. (1984) 365:1107–1113.

Godovac-Zimmermann J, Braunitzer G. The primary structure of alpha a- and beta-chains from Blue-and-yellow macaw (*Ara ararauna*, Psittaci) hemoglobin. No evidence for expression of alpha d-chains. Biol Chem Hoppe Seyler (1985) 366:503–508.

Goh SH, Lee YT, Bhanu NV, Cam MC, Desper R, Martin BM, Moharram R, Gherman RB, Miller JL. A newly discovered human alpha-globin gene. Blood (2005) 106:1466–1472.

Goldman N, Anderson JP, Rodrigo AG. Likelihood-based tests of topologies in phylogenetics. Syst Biol. (2000) 49:652–670.

Goodman M, Miyamoto MM, Czelusniak J. Pattern and process in vertebrate phylogeny revealed by coevolution of molecules and phylogenies. In: Molecules and morphology in evolution: conflict or compromise?—Patterson C, ed. (1987) New York: Cambridge University Press. 140–176.

Goodman M, Moore GW, Matsuda G. Darwinian evolution in the genealogy of haemoglobin. Nature (1975) 253:603–608.

Goodman M, Weiss ML, Czelusniak J. Molecular evolution above the species level: branching pattern rates and mechanisms. Syst Zool (1982) 31:376–399.

Gorr TA, Mable BK, Kleinschmidt T. Phylogenetic analysis of reptilian hemoglobins: trees rates and divergences. J Mol Evol. (1998) 47:471–485.

Grigg GC, Wells RMG, Beard LA. Allosteric control of oxygen binding by haemoglobin during embryonic development in the crocodile *Crocodylus porosus*: the role of red cell organic phosphates and carbon dioxide. J Exp Biol. (1993) 175:15–32.

Hall TA. Bioedit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucl Acids Symp Ser (1999) 41:95–98.

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

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

Hardison R. Organization, evolution, and regulation of the globin genes. In: Disorders of hemoglobin: genetics, pathophysiology, and clinical managment—Steinberg MH, Forget BG, Higgs DR, Nagel RL, eds. (2001) Cambridge: Cambridge University Press. 95–115.

Hiebl I, Braunitzer G, Schneeganss D. The primary structures of the major and minor hemoglobin-components of adult Andean goose (*Chloephaga melanoptera*: Anatidae): the muta-

tion Leu–--Ser in position 55 of the beta-chains. Biol Chem Hoppe Seyler (1987) 368:1559–1569.

Hiebl I, Schneeganss D, Braunitzer G. High-altitude respiration of birds. The primary structures of the alpha D-chains of the Barheaded goose (*Anser indicus*), the Greylag goose(A*nser anser*) and the Canada goose (*Branta canadensis*). Biol Chem Hoppe Seyler (1986) 367:591–599.

Hiebl I, Weber RE, Schneeganss D, Braunitzer G. High-altitude respiration of Falconiformes. The primary structures and functional properties of the major and minor hemoglobin components of the adult White-headed vulture (*Trigonoceps occipitalis*, Aegypiinae). Biol Chem Hoppe Seyler (1989) 370:699–706.

Hiebl I, Weber RE, Schneeganss D, Kosters J, Braunitzer G. High-altitude respiration of birds. Structural adaptations in the major and minor hemoglobin components of adult Ruppell's griffon (*Gyps rueppellii*, Aegypiinae): a new molecular pattern for hypoxic tolerance. Biol Chem Hoppe Seyler (1988) 369:217–232.

Higgs DR, Vickers MA, Wilkie AO, Pretorius IM, Jarman AP, Weatherall DJ. A review of the molecular genetics of the human alpha-globin gene cluster. Blood (1989) 73:1081–1104.

Hughes AL. Adaptive evolution of genes and genomes (1999) New York: Oxford University Press.

Hughes JR, Cheng J, Ventress N, Prabhakar S, Clark K, Anguita E, De Gobbi M, de Jong P, Rubin E, Higgs DR. Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences. Proc Natl Acad Sci USA (2005) 102:9830–9835.

Ikehara T, Eguchi Y, Kayo S, Takei H. 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 (1997) 234:450–453.

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

Kimura M, Ohta T. On some principles governing molecular evolution. Proc Natl Acad Sci USA (1974) 71:2848–2852.

Knapp JE, Oliveira MA, Xie Q, Ernst SR, Riggs AF, Hackert ML. The structural and functional analysis of the hemoglobin D component from chicken. J Biol Chem. (1999) 274:6411–6420.

Nagel RL, Steinberg MH. Role of epistatic (modifier) genes in the modulation of the phenotypic diversity of sickle cell anemia. Pediatr Pathol Mol Med (2001) 20:123–136.

Nothum R, Braunitzer G, Hiebl I, Kosters J, Schneeganss D. The hemoglobins of the adult blackbird (*Turdus merula*, Passeriformes). The sequence of the major (*HbA*) and minor component (*HbD*). Biol Chem Hoppe Seyler (1989) 370:309–316.

Nothum R, Weber RE, Kösters J, Schneeganss D, Braunitzer G. Amino-acid sequences and functional differentiation of hemoglobins A and D from swift (*Apus apus*, Apodiformes). Biol Chem Hoppe Seyler (1989) 370:1197–1207.

Ohno S. Evolution by gene duplication (1970) New York: Springer-Verlag.

Ohta T. Pattern of nucleotide substitutions in growth hormone-prolactin gene family: a paradigm for evolution by gene duplication. Genetics (1993) 134:1271–1276.

Ohta T. Further examples of evolution by gene duplication revealed through DNA sequence comparisons. Genetics (1994) 138:1331–1337.

Posada D, Crandall KA. Modeltest: testing the model of DNA substitution. Bioinformatics (1998) 14:817–818.

Proudfoot NJ, Gil A, Maniatis T. The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. Cell. (1982) 31:553–563.

Riggs AF. Self-association, cooperativity and supercooperativity of oxygen binding by hemoglobins. J Exp Biol. (1998) 201:1073–1084.

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

Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics (2003) 19:1572–1574.

Shimodaira H. An approximately unbiased test of phylogenetic tree selection. Syst Biol. (2002) 51:492–508.

Shimodaira H, Hasegawa M. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol. (1999) 16:1114–1116.

Shimodaira H, Hasegawa M. Consel: for assessing the confidence of phylogenetic tree selection. Bioinformatics (2001) 17:1246–1247.

Shishikura F. The primary structure of hemoglobin d from the Aldabra giant tortoise, *Geochelone gigantea*. Zoolog Sci. (2002) 19:197–206.

Stoeckelhuber M, Gorr T, Kleinschmidt T. The primary structure of three hemoglobin chains from the Indigo snake (*Drymarchon corais erebennus*, Serpentes): first evidence for alphaD chains and two beta chain types in snakes. Biol Chem. (2002) 383:1907–1916.

Swofford DL. Paup* phylogenetic analysis using parsimony (* and other methods) version 4.0b10 (2002) Sunderland, MA: Sinauer.

Swofford DL, Olsen GJ, Waddell PJ, Hillis DM. Phylogenetic inference. In: Molecular systematics 2nd edition—Hillis DM, Moritz C, Mable B, eds. (1996) Sunderland, MA: Sinauer. 407–514.

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

Thompson JD, Higgins DG, Gibson TJ. Clustal w: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. (1994) 22:4673–4680.

Weber RE, Hiebl I, Braunitzer G. High altitude and hemoglobin function in the vultures *Gyps rueppellii* and *Aegypius monachus*. Biol Chem Hoppe Seyler (1988) 369:233–240.

Wheeler D, Hope RM, Cooper SJB, Gooley AA, Holland RAB. 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. (2004) 58:642–652.

Yang Z. Paml: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci (1997) 13:555–556.

Zimmer EA, Martin SL, Beverley SM, Kan YW, Wilson AC. Rapid duplication and loss of genes coding for the alpha chains of hemoglobin. Proc Natl Acad Sci USA (1980) 77:2158–2162.

Supplemental Material

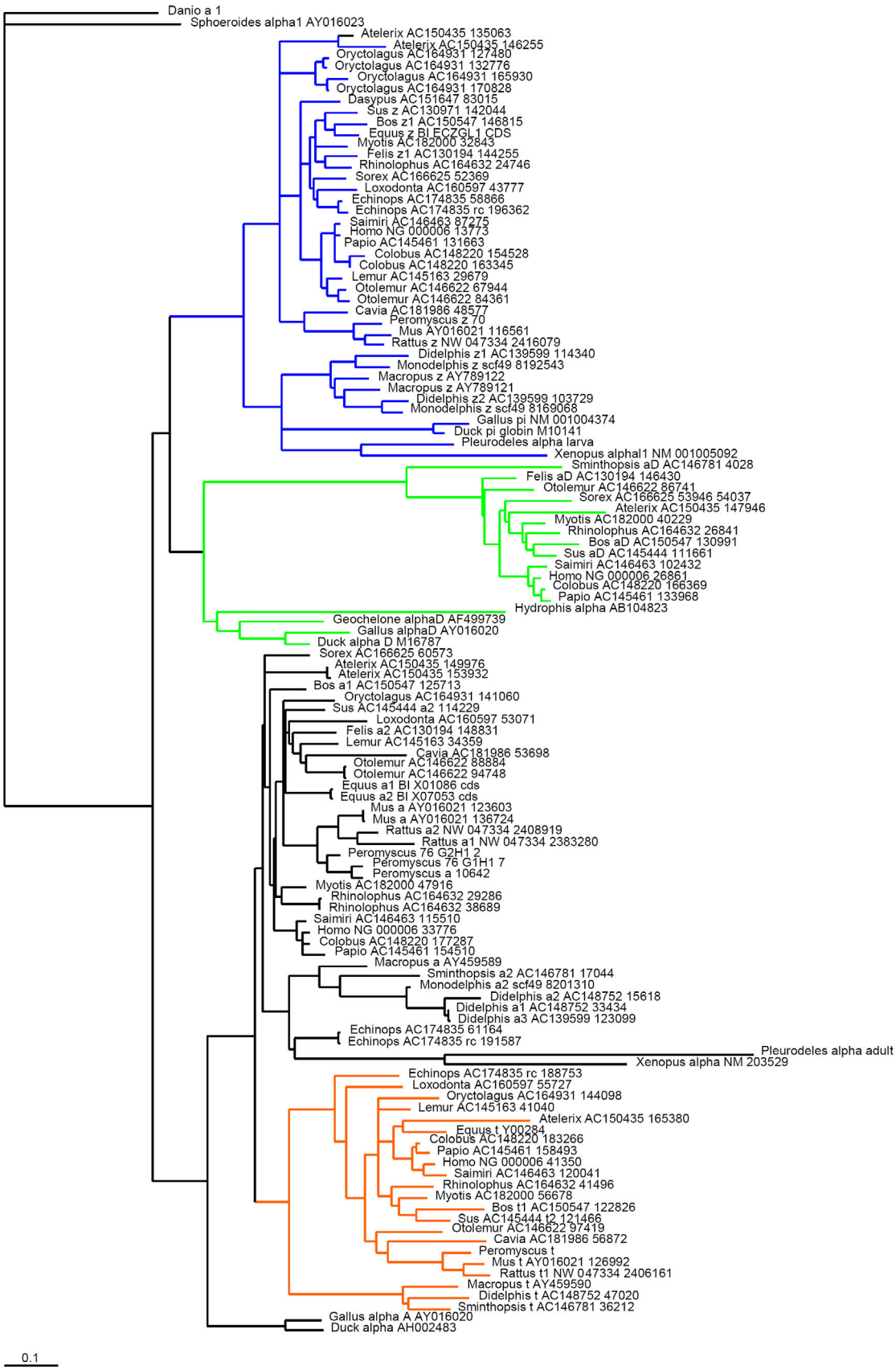

Figure S1: Phylogenetic relationships among the α-like globin genes of vertebrates estimated by neighbor-joining using a GTR+Γ model of nucleotide substitution. α^{E}-globin sequences (including amphibian α^{L}-globin, avian π-globin, and mammalian ζ-globin) are shown in blue, α^{D}-globin sequences are shown in green, α^{A}-globin sequences are shown in black, and θ-globin sequences are shown in red. Two fish α-globin sequences were used as outgroups.

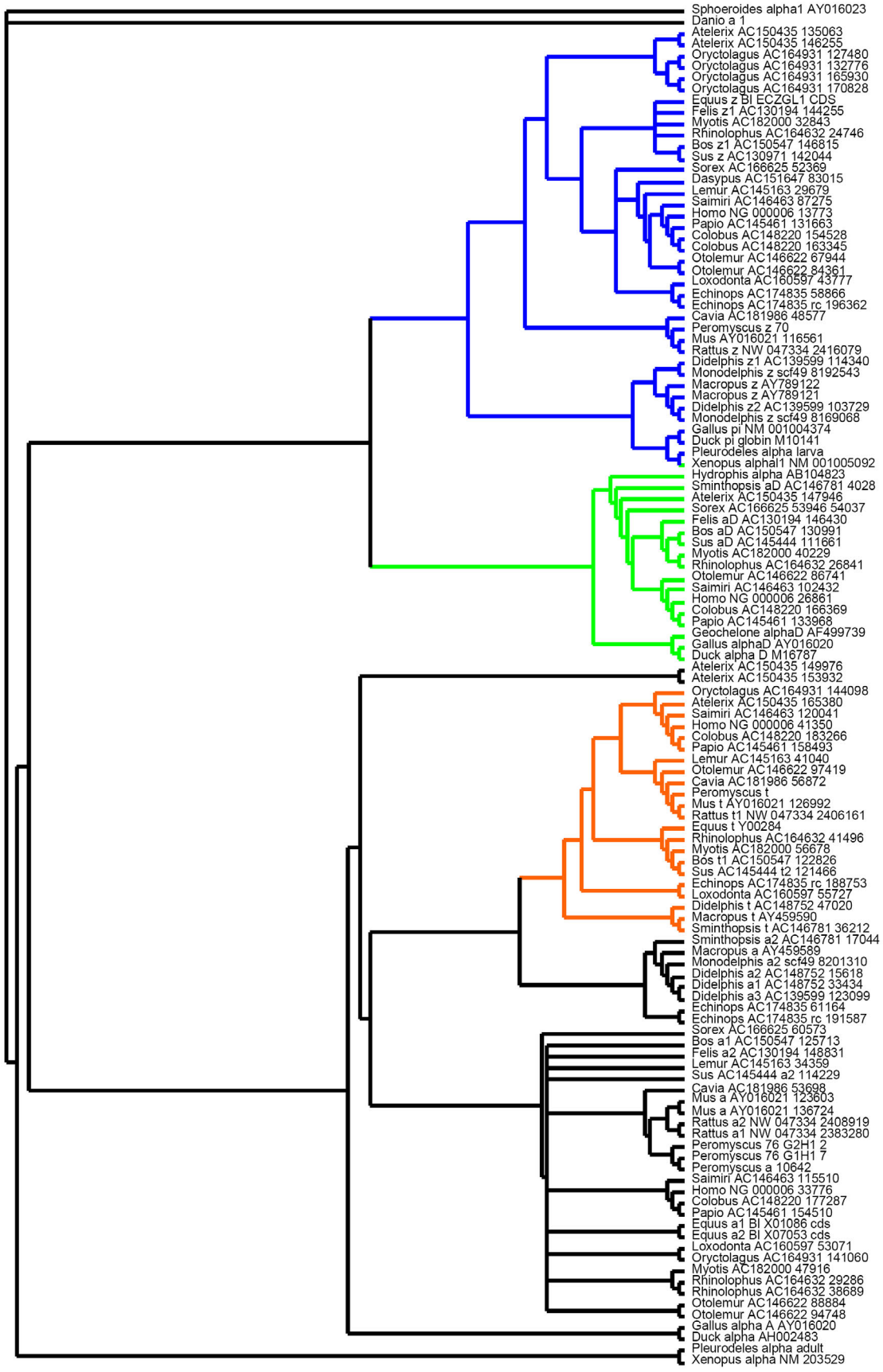

Figure S2: Phylogenetic relationships among the α-like globin genes of vertebrates estimated by parsimony. α^E-globin sequences (including amphibian α^L-globin, avian π-globin, and mammalian ζ-globin) are shown in blue, α^D-globin sequences are shown in green, α^A-globin sequences are shown in black, and θ-globin sequences are shown in red. Two fish α-globin sequences were used as outgroups

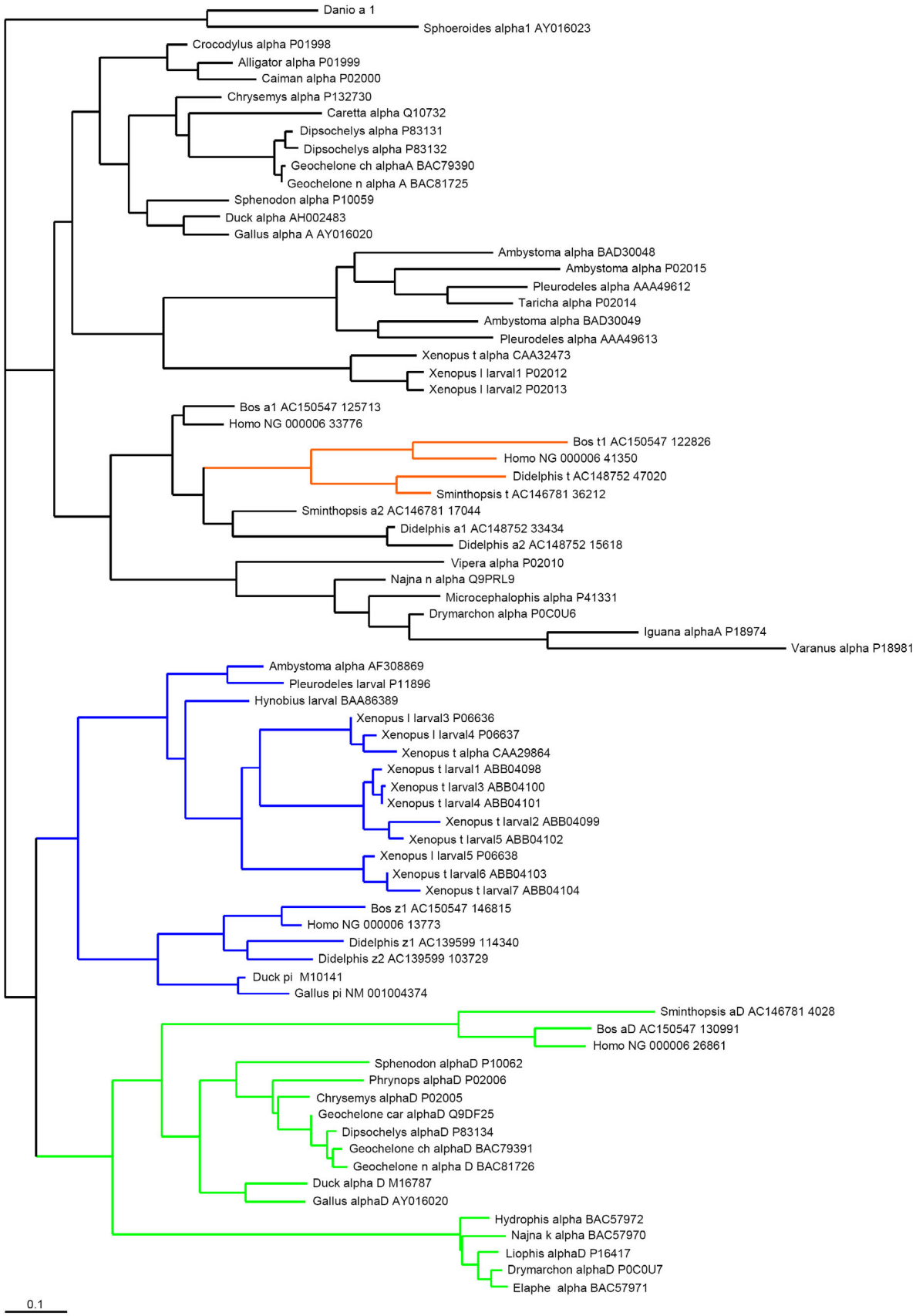

Figure S3: Maximum likelihood phylogram describing relationships among the α-like globin genes of vertebrates, with α^{E}-globin sequences (including amphibian α^{L}-globin, avian π-globin, and mammalian ζ-globin) are shown in blue, α^{D}-globin sequences are shown in green, α^{A}-globin sequences are shown in black, and θ-globin sequences are shown in red. Two fish α-globin sequences were used as outgroups. Searches were conducted in Treefinder version May 2006 (51) using an rtREV model of amino acid substitution.

Table S1. Summary of the DNA sequences used in this study. GC indicates that the sequences derive from a genomic contig.

Taxon ID	Common Name (used in Figure 1)	Accession Number M13365 (GenBank or ENSEMBL)	
EUTHERIAN MAMMALS			
AFROTHERIA			
Echinops telfairi	Tenrec	AC174835	GC
Loxodonta africana	Elephant	AC158446, AC160597	GC
XENARTHRA			
Dasypus novemcintus	Armadillo	AC151647	GC
LAURASIATHERIA			
Atelerix albiventris	Hedgehog	AC150435	GC
Bos taurus	Cow	AC150547	GC
Equus caballus	Horse	X01086, X07053, X07051, Y00284	
Felis catus	Domestic Cat	AC130194	GC
Myotis lucifugus	Little Brown Bat	AC182000	GC
Rhinolophus ferrumequinum	Horseshoe Bat	AC164632	GC
Sorex araneus	Shrew	AC166625	GC
Sus scrofa	Pig	AC145444, AC130971	GC
EUARCHONTOGLIRES			
Cavia porcellus	Guinea-pig	AC181986	GC
Colobus guereza	Colobus	AC148220	GC
Homo sapiens	Human	NG_000006	GC
Lemur catta	Lemur	AC145163	GC
Mus musculus	Mouse	AY016021	GC
Oryctolagus cuniculus	Rabbit	AC164931	GC
Otolemur garnetti	Galago	AC146622	GC
Papio hamadryas	Baboon	AC145461	GC
Peromyscus maniculatus	Deer Mouse		
Rattus norvegicus	Rat	NW_047334	GC
Saimiri boliviensis	Squirrel Monkey	AC146643	GC
MARSUPIALS			
Didelphis virginiana	Opossum	AC148752, AC139599	GC
Macropus eugenii	Tammar	AY459989, AY459590, AY789121, AY789122	
Monodelphis domestica	Short-tailed Opossum	scaffold_49 from MonDom 2.0 at ENSEMBL	GC
Sminthopsis macroura	Dunnart	AC146781	GC
BIRDS			
Gallus gallus	Chicken	AY016020, NM_001004374	GC
Cairina moschata	Duck	AH0024832, M10141, M16787	
REPTILES			
Geochelone denticulata	Tortoise	AF499739	
Hydrophis melanocephalus	Sea-snake	AB104823	
AMPHIBIANS			
Pleurodeles waltlii	Salamander	M13365, X14226	
Xenopus laevis	Frog	NM_203529, NM_001005092	
FISH			
Sphoeroides nephelus	Puffer fish	AY016023	
Danio rerio	Zebra fish	NM_131257	