Differential Loss and Retention of Cytoglobin, Myoglobin, and Globin-E during the Radiation of Vertebrates

Federico G. Hoffmann[1,2], Juan C. Opazo[3], and Jay F. Storz*[,1]

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

[2]Department of Biochemistry and Molecular Biology, Mississippi State University

[3]Instituto de Ecología y Evolución, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile

*Corresponding author: E-mail: jstorz2@unl.edu.

Accepted: 30 May 2011

Abstract

If rates of postduplication gene retention are positively correlated with levels of functional constraint, then gene duplicates that have been retained in a restricted number of taxonomic lineages would be expected to exhibit relatively low levels of sequence conservation. Paradoxical patterns are presented by gene duplicates that have been retained in a small number of taxa but which are nonetheless subject to strong purifying selection relative to paralogous members of the same multigene family. This pattern suggests that such genes may have been co-opted for novel, lineage-specific functions. One possible example involves the enigmatic globin-E gene (*GbE*), which appears to be exclusively restricted to birds. Available data indicate that this gene is expressed exclusively in the avian eye, but its physiological function remains a mystery. In contrast to the highly restricted phyletic distribution of *GbE*, the overwhelming majority of jawed vertebrates (gnathostomes) possess copies of the related cytoglobin (*Cygb*) and myoglobin (*Mb*) genes. The purpose of the present study was 1) to assess the phyletic distribution of the *Cygb*, *Mb*, and *GbE* genes among vertebrates, 2) to elucidate the duplicative origins and evolutionary histories of these three genes, and 3) to evaluate the relative levels of functional constraint of these genes based on comparative sequence analysis. To accomplish these objectives, we conducted a combined phylogenetic and comparative genomic analysis involving taxa that represent each of the major lineages of gnathostome vertebrates. Results of synteny comparisons and phylogenetic topology tests revealed that *GbE* is clearly not the product of a recent, bird-specific duplication event. Instead, *GbE* originated via duplication of a proto-*Mb* gene in the stem lineage of gnathostomes. Unlike the *Mb* gene, which has been retained in all major gnathostome lineages other than amphibians, the *GbE* gene has been retained only in the lineage leading to modern birds and has been independently lost in at least four major lineages: teleost fish, amphibians, mammals, and nonavian reptiles. Despite the restricted phyletic distribution of this gene, our results indicate that *GbE* is one of the most highly conserved globins in the avian genome.

Key words: bird genome, gene duplication, gene family evolution, genome evolution, globins.

Introduction

Duplicated genes are most likely to be retained in the genome when the two descendant copies partition ancestral subfunctions of the single-copy progenitor or when one or both copies evolve new specializations of function (Lynch and Force 2000; Lynch et al. 2001). Because different gene functions are not all equally essential in the economy of the organism, it is not surprising that paralogous members of the same multigene family often exhibit substantial variation in levels of functional constraint. Comparative analysis of genomic sequence data can be used to evaluate variation in functional constraint among members of the same gene family by measuring rates of gene loss and retention among different taxonomic lineages and by measuring rates of amino acid substitution in orthologous comparisons. Highly essential genes are expected to be retained in phylogenetically diverse taxa and are also expected to exhibit especially slow rates of amino acid substitution relative to paralogous members of the same gene family. At the other end of the spectrum, dispensable genes may be inactivated or deleted in multiple taxa, and such genes are also expected to exhibit elevated rates of amino acid substitution due to a relative

relaxation of purifying selection. Seemingly paradoxical patterns are presented by genes that have been retained in a small number of taxa but which nonetheless exhibit high levels of sequence conservation relative to other paralogous gene duplicates. This pattern suggests that such genes may have been co-opted for a novel, lineage-specific function.

Comparative genomic studies of the globin gene superfamily have revealed a number of striking discrepancies in the phyletic distribution of paralogous gene lineages. For example, all jawed vertebrates (gnathostomes) examined to date possess copies of neuroglobin (*Ngb*) and cytoglobin (*Cygb*; Burmester et al. 2000, 2002, 2004; Awenius et al. 2001; Trent and Hargrove 2002; Fuchs et al. 2004, 2005; Kugelstadt et al. 2004; Wystub et al. 2004; Hankeln et al. 2005; Roesner et al. 2005; Hankeln and Burmester 2008; Burmester and Hankeln 2009; Hoffmann, Opazo, and Storz 2010). These genes are also characterized by the highest levels of sequence conservation among all members of the gene family (Burmester et al. 2004). In gnathostomes, hemoglobin (Hb) and myoglobin (Mb) also represent examples of seemingly indispensable globin proteins that play critical roles in the maintenance of cellular oxygen supply in support of aerobic metabolism. Antarctic ice fish in the family Channichthyidae represent one notable exception to what appears to be an otherwise universal rule, as these fish do not express Hb and many species in the family do not express Mb either (Hemmingsen 1991; Sidell and O'Brien 2006). Mb has also been secondarily lost in amphibians (Maeda and Fitch 1982; Xi et al. 2007).

In contrast to the highly conserved Ngb, Cygb, Hb, and Mb proteins that have been retained in all or nearly all vertebrate lineages, there are also a number of paralogous globin proteins that have much more restricted phyletic distributions. The *globin X* gene is the product of a duplication event that predates the divergence between protostomes and deuterostomes, and among vertebrates, it has only been found in the genomes of teleost fish and *Xenopus* (Roesner et al. 2005). The *globin Y* (*GbY*) gene appears to be restricted to gnathostome vertebrates and has thus far only been found in the genomes of *Xenopus*, anole lizard, and platypus (Fuchs et al. 2006; Patel et al. 2008; Hoffmann, Opazo, and Storz 2010). The *GbY* gene and the proto-*Hb* gene represent paralogous products of an ancient tandem gene duplication in the stem lineage of vertebrates (Storz et al. 2011). The ancestral linkage arrangement of these two genes is still retained in the genomes of *Xenopus*, anole lizard, and platypus, as *GbY* is located at the 3′ end of the α-globin gene cluster in each of these taxa (Fuchs et al. 2006; Patel et al. 2008; Hoffmann, Storz, et al. 2010). Finally, the globin-E (*GbE*) gene has thus far only been found in the genome of birds (Kugelstadt et al. 2004; Hoffmann, Opazo, and Storz 2010). Available data suggest that *GbE* is expressed exclusively in the eye, and although structural considerations suggest that the protein could have an oxygen-binding respiratory function, its physiological role remains a mystery (Kugelstadt et al. 2004). The level of functional constraint of this gene has not been assessed previously because it has been characterized only in chicken (Kugelstadt et al. 2004). Thus, it has not been possible to make the same kinds of taxonomic cross-comparisons that have been made with other globin genes that have a broader phyletic distribution. However, with the availability of genomic sequence data from other bird species besides chicken (turkey, mallard duck, and zebra finch), it is now possible to compare levels of among-species sequence conservation across the full complement of globin genes in the avian genome.

There have been conflicting accounts concerning the phylogenetic placement of *GbE* relative to other vertebrate-specific globin genes. Based on results of a phylogenetic analysis, which indicated that *GbE* is most closely related to *Cygb*, Kugelstadt et al. (2004) suggested two alternative hypotheses: 1) *GbE* and *Cygb* trace their duplicative origins to the common ancestor of tetrapods and teleost fish, in which case the *GbE* gene would have been secondarily lost from all descendant lineages other than birds or 2) *GbE* is the product of a bird-specific duplication event. The phylogenetic reconstructions reported by Hoffmann, Opazo, and Storz (2010) are not consistent with either of these hypotheses. Instead, these more recent results indicate that *GbE* originated via duplication of a proto-*Mb* gene in the stem lineage of gnathostomes. The latter scenario also posits that the *GbE* gene was lost secondarily from all descendant lineages other than birds.

The purpose of the present study was 1) to assess the phyletic distribution of the *Cygb*, *Mb*, and *GbE* genes among vertebrates, 2) to elucidate the duplicative origins and evolutionary histories of these three genes, and 3) to evaluate the relative levels of functional constraint of these genes based on comparative sequence analysis. To accomplish these objectives, we conducted a combined phylogenetic and comparative genomic analysis involving taxa that represent each of the major lineages of gnathostome vertebrates. Results of synteny comparisons and phylogenetic topology tests revealed that the *Mb* and *GbE* genes represent products of a tandem gene duplication in the stem lineage of gnathostomes and that the proto-*Mb*/*GbE* gene and the *Cygb* gene represent paralogous products of a whole-genome duplication event in the stem lineage of vertebrates. Thus, *GbE* is clearly not the product of a recent, bird-specific duplication event. Rather, available data suggest that this ancient gene has been retained only in birds (and possibly in other archosaur lineages) and has been independently lost in all other gnathostome lineages. Despite the restricted phyletic distribution of this gene, our results also indicate that, in birds, *GbE* is one of the most highly conserved globin proteins and is characterized by a level of functional constraint similar to that of both *Mb* and *Hb*.

Materials and Methods

Data Collection

We had previously identified the full complement of globin genes in the genomes of 12 vertebrate species (Hoffmann, Opazo, and Storz 2010): five teleost fish (fugu, *Takifugu rubripes*; medaka, *Oryzias latipes*; pufferfish, *Tetraodon nigroviridis*; three-spined stickleback, *Gasterosteus aculeatus*; and zebra fish, *Danio rerio*), one amphibian (western clawed frog, *Xenopus tropicalis*), one squamate reptile (green anole lizard, *Anolis carolinensis*), two birds (chicken, *Gallus gallus* and zebra finch, *Taeniopygia guttata*), and three mammals (human, *Homo sapiens*; gray short-tailed opossum, *Monodelphis domestica*; and platypus, *Ornithorhynchus anatinus*). For the present study, we added sequences corresponding to the *Cygb*, *GbE*, and *Mb* genes from the prerelease version of the mallard duck (*Anas platyrhynchos*) and turkey (*Meleagris gallopavo*) genomes available from the Ensembl-Pre database. For the purpose of investigating patterns of conserved synteny in mammals, we also annotated the genes found upstream and downstream of the *Mb* gene in cow (*Bos taurus*), dog (*Canis familiaris*), rabbit (*Oryctolagus cuniculus*), guinea pig (*Cavia porcellus*), mouse (*Mus musculus*), and rat (*Rattus norvegicus*). The additional globin sequences were annotated by comparing known coding sequences to genomic contigs using basic alignment search tool (Blast) (Altschul et al. 1990).

Phylogenetic Analyses

To verify the phylogenetic position of the additional avian globin sequences included in this study, we added sequences from the *Cygb*, *GbE*, and *Mb* genes of turkey and mallard duck to an alignment of vertebrate globins that included representative amino acid sequences of *Cygb* (including cyclostome *Hbs*; Hoffmann, Opazo, and Storz 2010), *GbE*, *GbY*, α- and β-chain *Hbs*, and *Mb*. As outgroup sequences, we included three globin sequences from amphioxus (*Branchiostoma floridae*, Cephalochordata), which are sister to the complete set of vertebrate-specific globins (Ebner et al. 2010; Storz et al. 2011), plus *Ngb* sequences from eight vertebrate taxa (supplementary table 1, Supplementary Material online). We aligned amino acid sequences using the L-INS-i strategy from Mafft version 6.8 (Katoh and Toh 2008; Katoh et al. 2009), and we estimated phylogenetic relationships using maximum likelihood and Bayesian approaches (the resulting alignment is provided as supplementary fig. l, Supplementary Material online). Maximum likelihood searches were carried out using Treefinder version October 2008 (Jobb et al. 2004) using the LG + I + Γ model (Le and Gascuel 2008), which was identified as the best fitting model of amino acid substitution with the "propose model" subroutine from Treefinder. Support for the nodes was evaluated with 1,000 bootstrap pseudoreplicates. Bayesian analyses were conducted in MrBayes v.3.1.2 (Ronquist and Huelsenbeck 2003) running six simultaneous chains for 10,000,000 generations, sampling every 1,000 generations, under a mixed model of amino acid substitution 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. MrBayes analyses were run using the Cipres portal (Miller et al. 2009).

Synteny Analyses

To assess synteny conservation among chromosomal regions that harbor the *Cygb*, *Mb*, and *GbE* genes, we annotated the genes found upstream and downstream of the paralogous globin genes in each of the 14 vertebrate genome sequences examined. Initial ortholog predictions were derived from EnsemblCompara (Vilella et al. 2009). We also used the program Genscan (Burge and Karlin 1997) to identify additional unannotated genes lying upstream and downstream of the annotated globin genes. The unannotated genes were compared with the nonredundant protein database using Blast (Altschul et al. 1990). Finally, partial sequences for genes of interest (representing either pseudogenes or incomplete coverage) were identified and annotated with Blast. Additional synteny comparisons were conducted using Genomicus (Muffato et al. 2010).

Hypothesis Testing

To test alternative phylogenetic hypotheses about pathways of globin gene family evolution, we used parametric bootstrapping tests (Swofford et al. 1996; Goldman et al. 2000) as well as the approximately unbiased (AU) test proposed by Shimodaira (2002). Results of Kugelstadt et al. (2004) suggested that the *GbE* and *Cygb* genes are either products of a relatively ancient duplication event in the gnathostome lineage that occurred prior to the split between tetrapods and teleost fish or they represent the products of a much more recent bird-specific duplication. By contrast, the phylogenetic reconstruction reported by Hoffmann, Opazo, and Storz (2010) suggests that *GbE* originated via duplication of a proto-*Mb* gene in the stem lineage of gnathostomes. To test these alternative hypotheses about the duplicative origins of the *Cygb*, *Mb*, and *GbE* genes, we compared the likelihood scores associated with the following phylogenies: 1) a constrained tree in which *GbE* was sister to the gnathostome *Cygb* clade, 2) a constrained tree in which *GbE* was sister to the avian *Cygb* clade, 3) a constrained tree in which *GbE* was sister to the avian *Mb* clade, and 4) an unconstrained tree in which *GbE* was sister to the clade that included the *Mb* genes of gnathostomes. All tests were carried out in Treefinder version October 2008 (Jobb et al. 2004).

Molecular Evolution Analyses

To assess relative levels of functional constraint among the different globin paralogs and to evaluate the effects of substitution rate heterogeneity on the phylogenetic reconstructions, we compared rates of amino acid substitution among avian orthologs of *Cygb*, *GbE*, and *Mb* using Tajima's (1993) relative rate test, as implemented in MEGA version 5 (Kumar et al. 2008).

To make more refined inferences about levels of sequence conservation and functional constraint among avian globin genes, we quantified sequence conservation among all members of the globin gene superfamily in the avian genome: *Cygb*, *Mb*, *GbE*, *Ngb*, α^E-globin, α^D-globin, α^A-globin, ρ-globin, ϵ-globin, β^A-globin, and β^H-globin (supplementary table 2, Supplementary Material online). Of the four bird species included in this analysis (all of which are representatives of the supraordinal clade Neognathae), the deepest split is between Neoaves (represented by zebra finch [order Passeriformes]) and Galloanserae (represented by chicken and turkey [order Galliformes] and mallard duck [order Anseriformes]). The Galloanserae–Neoaves divergence is thought to have occurred ~100–130 Ma in the Cretaceous (Ericson et al. 2006; Brown et al. 2008). For comparative purposes, we conducted a similar analysis of sequence conservation among mammalian globin genes. To do this, we chose a phylogenetic assemblage of mammalian taxa that are thought to have diversified over roughly the same timescale as the four bird species that we examined. Specifically, we analyzed the evolution of globin genes in representative species from the supraordinal clade Boreoeutheria, which contains Lausariatheria (the superorder that includes bats, carnivores, cetaceans, artiodactyls, and perissodactyls) and Euarchontoglires (the superorder that includes primates, rodents, and lagomorphs). This clade provides a good basis for comparison with the neognathe birds, as the split between Laurasiatheria and Euarchontoglires is thought to have occurred roughly 100 Ma in the mid-Cretaceous (Bininda-Emonds et al. 2007; Murphy et al. 2007). We compared relative levels of sequence conservation among all members of the globin gene superfamily shared by boreoeutherian mammals: *Cygb*, *Mb*, *Ngb*, α^E-globin (= ζ-globin), α^D-globin, α^A-globin, ϵ-globin, and β-globin. This analysis included the full complement of globin genes found in the genomes of mega- and microchiropteran bats, cow, dog, guinea pig, mouse, pika, rabbit, human, macaque, and marmoset (supplementary table 2, Supplementary Material online).

Nucleotide sequences were retrieved from the GenBank or Ensembl databases, and conceptual translations of the nucleotide sequences were then aligned using the L-INS-i strategy in Mafft version 6.8 (Katoh and Toh 2008; Katoh et al. 2009). The amino acid alignments were then back translated to using the Mobyle web server, and the resulting nucleotide alignments were used in all subsequent analyses of molecular evolution. The alignment of globin sequences for neognathe birds is available as supplementary figure 2 (Supplementary Material online) and the alignment of globin sequences for boreoeutherian mammals is available as supplementary figure 3 (Supplementary Material online). In separate analyses of the avian and mammalian globin genes, we calculated the average pairwise amino acid differences using MEGA version 5 (Kumar et al. 2008), and we compared estimates of ω ($=d_N/d_S$), the ratio of the rate of nonsynonymous substitution per nonsynonymous site (d_N) to the rate of synonymous substitution per synonymous site (d_S) using the codeml program from the PAML package version 4.4c (Yang 2007).

Results and Discussion

Phyletic Distribution

To unravel the evolutionary history of the *Cygb*, *Mb*, and *GbE* gene lineages, we first assessed the phyletic distribution of the three paralogs among representatives of each of the major vertebrate classes for which complete genomic sequence was available: teleost fish, amphibians, mammals, birds, and nonavian reptiles. All tetrapod species examined in this study possess a single copy of the *Cygb* gene, and all teleost fish possess either two duplicated copies or, in the case of zebra fish, three copies (fig. 1). Analysis of conserved synteny revealed that the two *Cygb* paralogs possessed by all teleost fish, *Cygb-1* and *Cygb-2*, are embedded in chromosomal segments that share multiple paralogous gene duplicates in the upstream and downstream flanking regions. These homologous chromosomal segments (or "paralogons") appear to represent products of a large-scale segmental duplication or possibly the whole-genome duplication event that occurred in the stem lineage of teleost fish (Meyer and Schartl 1999; Taylor et al. 2001, 2003). In fact, we were able to confirm that the *Cygb-1* and *Cygb-2* paralogs derive from the fish-specific whole-genome duplication because both *Cygb*-defined paralogons descend from chromosome "*e*" in the reconstructed (unduplicated) protokaryotype of the teleost common ancestor (Nakatani et al. 2007). This supports the conjecture of Fuchs et al. (2005) that the *Cygb-1* and *Cygb-2* paralogs are products of whole-genome duplication.

Almost all species examined possess a single copy of the *Mb* gene. Consistent with previous reports (e.g., Fuchs et al. 2006; Xi et al. 2007), we found no trace of an *Mb* gene in the *Xenopus* genome. The absence of *Mb* in amphibians in the order Anura, which includes frogs and toads, and the order Caudata, which includes salamanders and newts, was previously noted in a survey of protein sequence data in the American bullfrog (*Rana catesbeiana*; Maeda and Fitch 1982), a survey of genomic sequence data in the western long-clawed frog (*X. tropicalis*; Fuchs et al. 2006), and

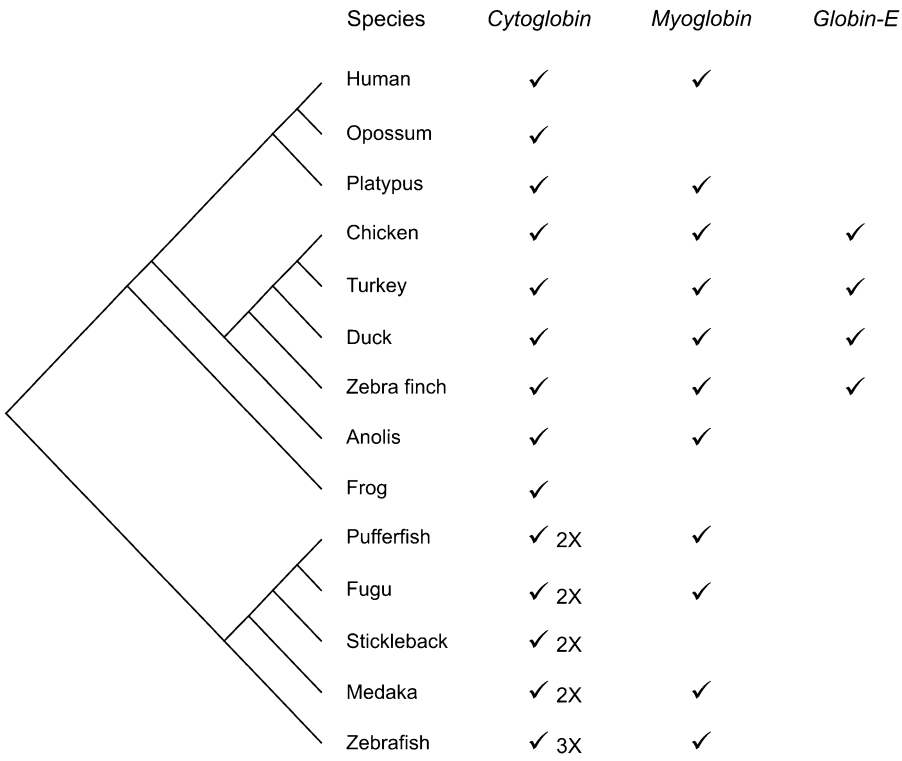

Fig. 1.—Phyletic distribution of *Cygb*, *GbE*, and *Mb* in the genomes of gnathostome vertebrates. Phylogenetic relationships among vertebrates were drawn following a consensus from Meyer and Zardoya (2003) and Li et al. (2007).

southern hybridization and reverse transcriptase-polymerase chain reaction experiments in Japanese fire-bellied newts (*Cynops pyrrhogaster*; Xi et al. 2007). Surprisingly, we found suggestive evidence that the *Mb* gene may have been wholly or partially deleted in two other vertebrate species, the three-spined stickleback and the gray short-tailed opossum. We found no trace of the *Mb* gene in the genome assembly of the three-spined stickleback, and we found only a partial coding sequence (exon 1) in the genome assembly of the opossum. We also searched the available stickleback and opossum expressed sequence tag (EST) databases and trace archives and found no *Mb* sequence matches.

Whereas one or more copies of the *Cygb* gene was found in all species examined and a single copy of the *Mb* gene was found in almost all species examined, an ortholog of the chicken *GbE* gene was only found in the genomes of the three additional avian species included in this study (turkey, mallard duck, and zebra finch; fig. 1). These avian *GbE* genes had the three exons/two introns structure typical of most vertebrate globin genes and ranged in length from 3,426 bp in duck to 4,064 bp in turkey (supplementary table 3, Supplementary Material online). The three exons were 95, 232, and 129 bp long, and the introns were located in the canonical B12.2. and G7.0 positions, as previously reported for the chicken *GbE* gene (Kugelstadt et al. 2004). Alignment of avian *GbE* amino acid sequences to human *Mb* reveals a similar distribution of hydrophobic and hydrophilic residues. Moreover, this alignment reveals that functionally important residues, such as the CD1 phenylalanine, the proximal (F8) histidine, and the distal (E7) histidine are conserved in all *GbE* sequences (fig. 2). Bioinformatic searches for previously unidentified orthologs of the avian *GbE* gene in vertebrate sequence databases failed to detect any significant matches in the genomes of nonavian taxa. These searches were based on both nucleotide and amino acid sequences and included EST and trace records from cyclostomes, cartilaginous fish, teleost fish, and tetrapods. It remains to be seen whether *GbE* is present in the genomes of other extant archosauromorphs (such as crocodilians or tuatara) or whether it is exclusively restricted to modern birds.

Phylogenetic Analyses

Results of maximum likelihood and Bayesian phylogenetic searches yielded a tree that is consistent with previous phylogenetic reconstructions of vertebrate-specific globin genes (Hoffmann, Opazo, and Storz 2010; Storz et al. 2011). The different vertebrate-specific globin paralogs included in this study formed four distinct and strongly supported clades: 1) *Cygb* + cyclostome *Hb*, 2) *Mb* + *GbE*, 3) *GbY*, and 4) the α- and β-chain *Hbs* of gnathostomes (fig. 3). Support for the monophyly of each of these four clades was strong in Bayesian analyses but moderate in the case of the maximum likelihood bootstrap analysis. In all analyses, we confirmed a sister relationship between the (*Cygb* +

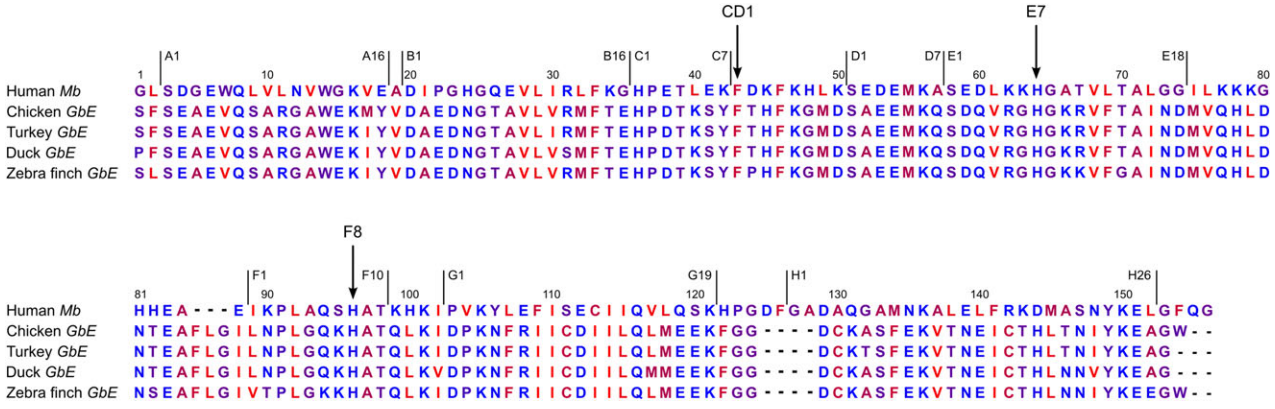

Fig. 2.—Structural alignment of avian GbE sequences to Human Mb. Amino acids are colored according to the hydrophobicity table of Kyte and Doolittle (1982), as implemented in JalView version 2.6.1 (Waterhouse et al. 2009). The most hydrophobic residues are shown in red, the most hydrophilic residues are shown in blue, and intermediate residues are shown in shades of purple. The α-helical structure is shown on top of the alignments. The functionally important phenylalanine in position CD1 and the distal and proximal histidines in positions E7 and F8 are indicated by arrows.

cyclostome *Hb*) and the (*Mb* + *GbE*) clades, whereas the deepest split among the four lineages of vertebrate-specific globins corresponded to the split between the α- and β-*Hbs* of gnathostomes and the common ancestor of the remaining three clades of paralogous globins. Again, only Bayesian analyses provided relatively high levels of support for these different clades. As shown in figure 3, the avian *Cygb*, *Mb*, and *GbE* sequences each form monophyletic groups, and in the case of *Cygb* and *Mb*, they are nested within the expected set of orthologous genes from other tetrapods.

The phylogenetic reconstruction shown in figure 3 is not consistent with that reported by Kugelstadt et al. (2004) (see also Burmester et al. 2004; Burmester and Hankeln 2009), probably due to the more dense sampling of vertebrate globin diversity made possible by the increased availability of genomic data. This previous study suggested that *GbE* was more closely related to *Cygb* than to other vertebrate-specific globins. Accordingly, Kugelstadt et al. (2004) suggested two alternative hypotheses: 1) *GbE* and *Cygb* are products of an ancient gene duplication that occurred prior to the split between teleost fish and tetrapods or 2) the *GbE* gene arose via duplication of the *Cygb* gene in the common ancestor of birds. With respect to the latter hypothesis, Kugelstadt et al. (2004) suggested that the observed placement of the avian *GbE* gene outside the clade of tetrapod *Cygb* sequences might be an artifactual result that is attributable to an accelerated rate of amino acid substitution.

We conducted topology tests to evaluate alternative phylogenetic hypotheses regarding the duplicative origins of *GbE* (fig. 4). Topology tests implemented with constrained trees strongly rejected the bird-specific duplication hypotheses in favor of the tree that placed *GbE* sister to the *Mb* clade (table 1). Moreover, the parametric bootstrap tests strongly rejected the ((*GbE*, *Cygb*) *Mb*) topology in favor of the ((*GbE*, *Mb*) *Cygb*) topology. The AU test also favored this latter topology although the difference in likelihood scores was only marginally significant ($P < 0.08$). Because our sample of *Mb* sequences included cartilaginous fish, teleost fish, and tetrapods, our results indicate that the tandem gene duplication that gave rise to the *GbE* and *Mb* genes occurred prior to the radiation of gnathostome vertebrates. These results also indicate that orthologs of the *GbE* gene have been lost independently in all gnathostome lineages other than the lineage leading to modern birds. The differential retention of *Mb* and *GbE* mirrors a similar pattern of gene turnover in the α and β globin gene clusters of tetrapod vertebrates (Hoffmann et al. 2008a,b; Opazo et al. 2008a,b).

Synteny Analyses

Because of the additional round of whole-genome duplication in teleost fish and the ensuing differential loss of paralogous gene duplicates, we initially conducted separate analyses on genomic data from tetrapods and teleost fish. In the case of the *Cygb* gene region, the analysis revealed a high degree of conserved synteny within tetrapods and within teleost fish (with the exception of zebra fish; fig. 5). Furthermore, the genes found downstream of the *Cygb-1* paralog in medaka, pufferfish, and stickleback are nearly identical to the genes found downstream of the *Cygb* gene in all tetrapods, although the linear order has been inverted (fig. 5). Similarly, the *SPHK1* and *PRSAP1* genes, which bracket the *Cygb-2* paralog in teleost fish, are also found in close proximity to the single-copy *Cygb* gene of tetrapods.

The pattern of conserved synteny spanning the *GbE* and *Mb* genes is more complex than that observed for *Cygb*. The genes found adjacent to *Mb* are well conserved within tetrapods (fig. 6*A*) and within teleost fish (fig. 6*B*). Among teleost fish, the zebra fish genome is again characterized by the lowest level of conserved synteny (fig. 6*B*). Synteny conservation

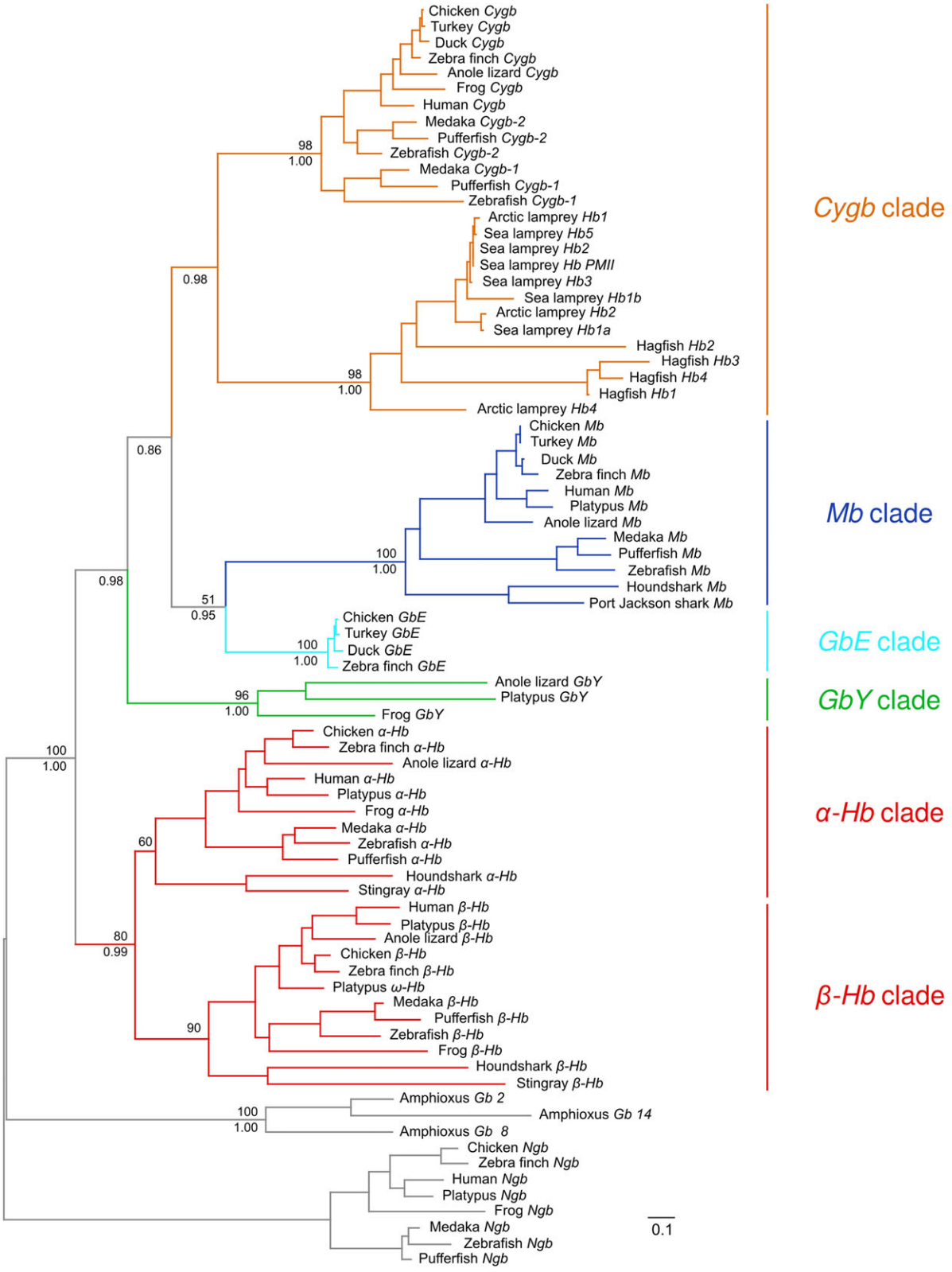

Fig. 3.—Maximum likelihood phylogram describing phylogenetic relationships among vertebrate-specific globins from representative taxa obtained under the LG + I + Γ model of amino acid substitution. As outgroup sequences, we included three globin sequences from amphioxus that are sister to the complete set of vertebrate-specific globins, plus *Ngb* sequences from eight representative vertebrate taxa. Numbers above the nodes correspond to maximum likelihood bootstrap support values, and those below the nodes correspond to Bayesian posterior probabilities.

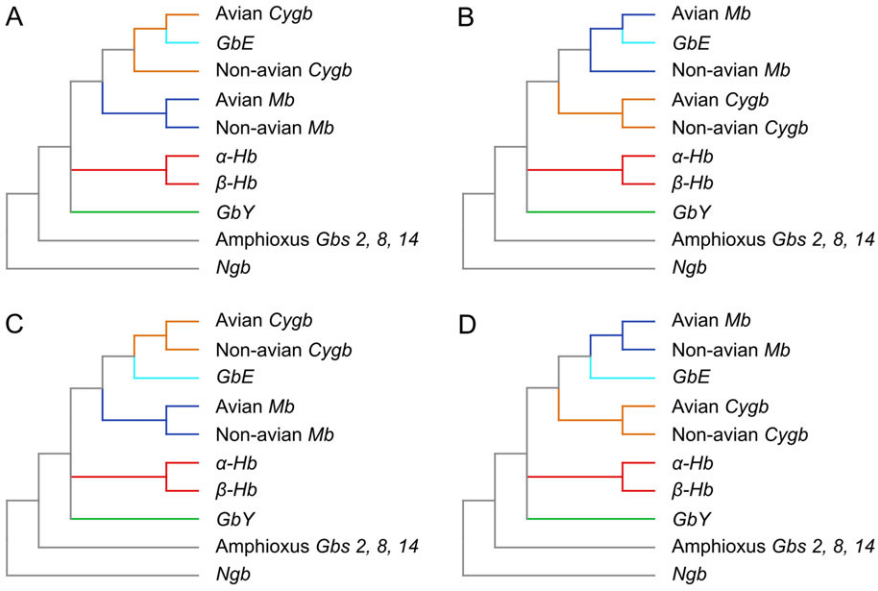

Fig. 4.—Schematic representation of alternative hypotheses regarding the phylogenetic position of *GbE*. According to the scenario shown in panel (*A*), *GbE* represents the product of a bird-specific duplication of the *Cygb* gene. According to the scenario shown in panel (*B*), *GbE* represents the product of a bird-specific duplication of the *Mb* gene. According to the scenario shown in panel (*C*), *GbE* originated via duplication of the proto-*Cygb* gene in the stem lineage of gnathostomes. According to the scenario shown in panel (*D*), *GbE* originated via duplication of the proto-*Mb* gene in the stem lineage of gnathostomes. Our phylogenetic and synteny analyses support hypothesis D.

is stronger in the case of amniotes, as represented by human, opossum, platypus, anole lizard, chicken, and zebra finch. If we ignore variation in the number of tandemly linked *APO* paralogs, the set of genes found upstream and downstream of the *Mb* gene are almost identical in this set of taxa (fig. 6*A*). Despite the fact that the stickleback and opossum appear to be missing orthologs of the *Mb* gene, synteny is conserved for the genes flanking the ancestral *Mb* locus in both species (fig. 6*A* and *B*). Contrary to the highly conserved pattern shown in figure 6, Xi et al. (2007) reported a low degree of conserved synteny in the *Mb* gene region of vertebrates. The discrepancy between our results and those of Xi et al. (2007) is attributable to the fact that these authors included mouse as a representative mammal in their synteny comparisons. Due to a chromosomal translocation that occurred in the common ancestor of rat and mouse, the *RASD2*, *MCM5*, *HMOX1*, *TOM1*, and *ISX* genes of these two rodent species are not found immediately adjacent to the *Mb* gene as they are in most other amniote taxa (supplementary fig. 4, Supplementary Material online).

Table 1
Likelihood Scores Associated with Alternative Phylogenetic Hypotheses Regarding the Relationships of *Cygb*, *Mb*, and *GbE* and Results from the Corresponding Topology Tests

Hypothesis	log-likelihood	AU test	Parametric Bootstrap Test
GbE sister to *Mb*	−14,989.86		
GbE sister to *Cygb*	−15,000.49	$P < 0.08$	$P < 0.001$
GbE sister to bird *Cygb* clade	−15,044.23	$P < 0.001$	$P < 0.001$
GbE sister to bird *Mb* clade	−15,064.81	$P < 0.001$	$P < 0.001$

In each of the examined avian genomes, the *GbE* gene is located between the *Luc7L2* and the *POLDIP3* genes (fig. 6*A*). In the anole lizard genome assembly, the *Luc7L2* and *POLDIP3* genes are located 5 kb apart on the same scaffold, but there is no trace of *GbE* in the intervening chromosomal region. In placental mammals, the *Luc7L2* and *POLDIP3* genes are located on separate chromosomes, whereas in opossum, they are separated by over 17 Mb on chromosome 8. Although the avian *Mb* and *GbE* genes are separated by physical distances of over 3 Mb in each of the bird species examined here, *Mb*- and *GbE*-associated genes that are separated by vast physical distances on chromosome 1 of galliform birds (and chromosome 1A of zebra finch) are found in close physical proximity in the homologous chromosomal segments of teleost fish and amphibians (fig. 6*A* and *B*; supplementary fig. 5, Supplementary Material online). This indicates that the *Mb* and *GbE* genes were closely linked in the genome of the gnathostome common ancestor.

Patterns of Molecular Evolution

In chicken and zebra finch, we did not detect significant differences in relative rates of amino acid substitution in the comparison between *Mb* and *GbE*. In turkey and mallard duck, by contrast, test results indicated that the *GbE* gene is evolving at a significantly slower rate than the *Mb* gene (table 2). The discrepancy in substitution rates between the

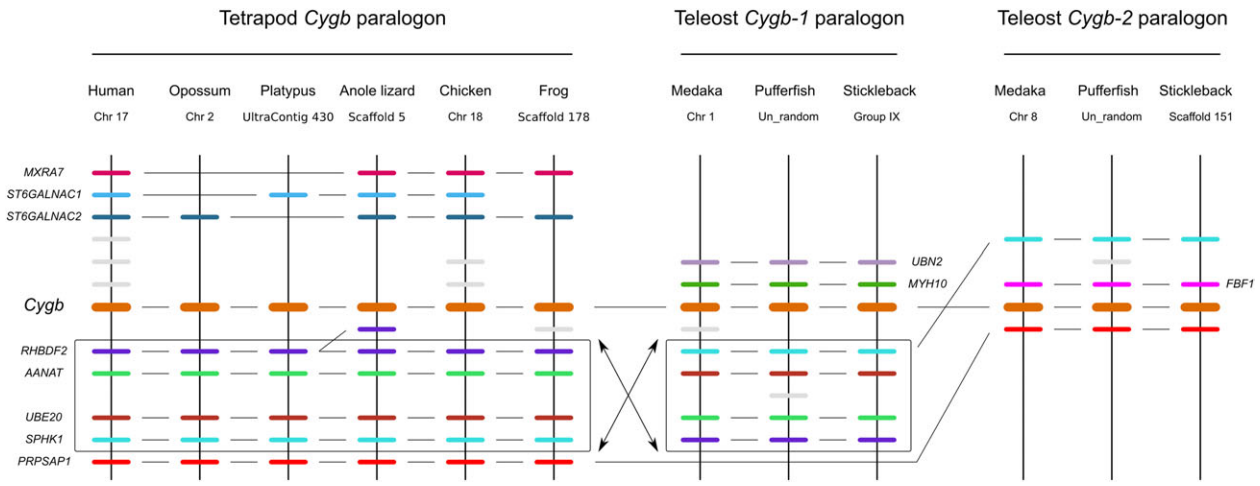

Fig. 5.—Patterns of conserved synteny in the chromosomal region that harbors the *Cygb* gene in gnathostome vertebrates. Horizontal lines denote orthologous relationships.

Mb and the *GbE* paralogs is also reflected in the reconstructed phylogeny, as the branch leading to the *GbE* clade is clearly shorter than the branch leading to the *Mb* clade (fig. 3).

All gnathostome vertebrates examined to date possess one or more copies of the *Cygb* gene, and the overwhelming majority possess a single *Mb* gene. By contrast, the *GbE* gene is restricted to birds and possibly other archosaurs. If rates of gene retention are positively correlated with the degree of functional constraint, then we would expect to observe the following rank order of sequence conservation: *Cygb*

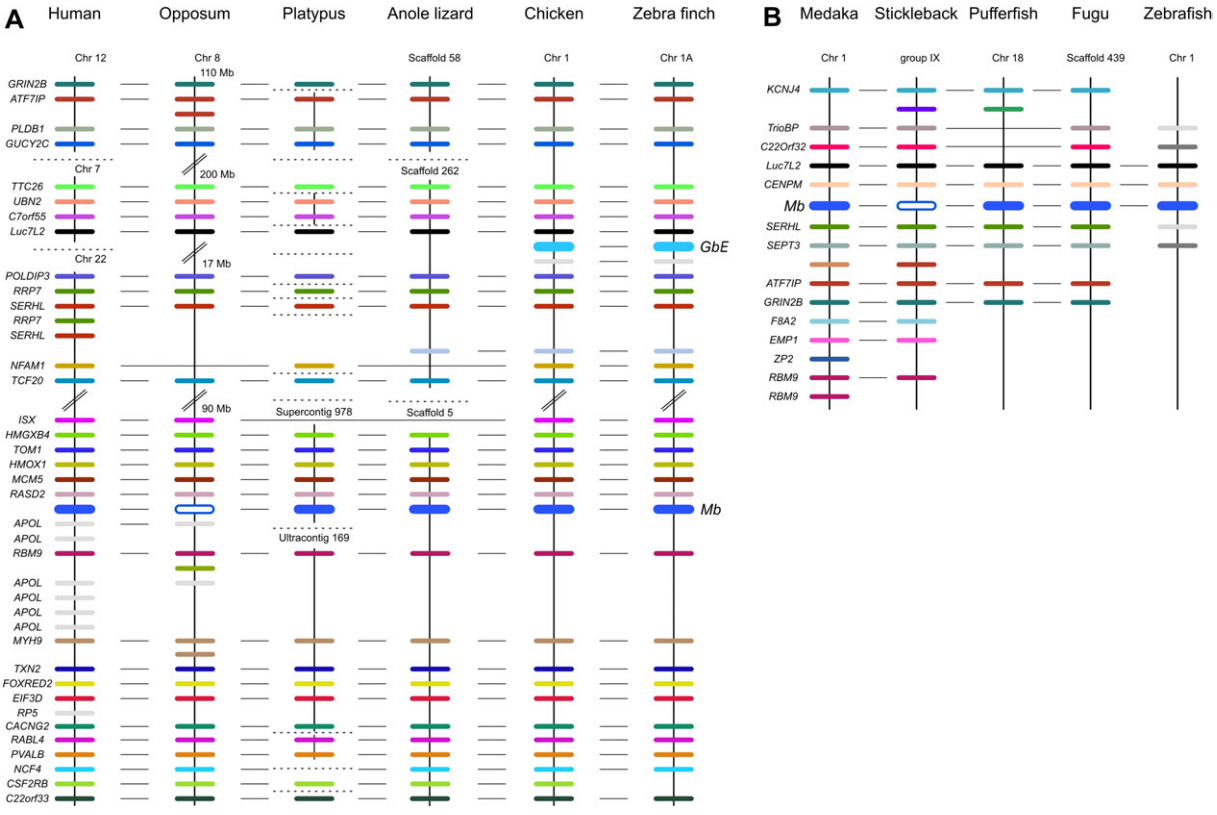

Fig. 6.—Patterns of conserved synteny in the chromosomal region that harbors the *GbE* and *Mb* genes in gnathostome vertebrates. (*A*) Synteny comparisons of the *GbE* and *Mb* gene regions among representative amniote taxa. (*B*) Synteny comparisons of the *Mb* gene region among representative teleost fish. Horizontal lines denote orthologous relationships.

Table 2
Results from Tajima's (1993) Relative Rate Tests of Amino Acid (aa) Substitution

	Unique aa Changes in *Mb*	Unique aa Changes in *GbE*	Unique aa Changes in *Cygb*	Identical Sites	Divergent Sites[a]	*P* Value
Chicken	27	15	21	27	57	NS
Duck	27	13	20	27	60	$P < 0.03$
Turkey	28	15	20	27	57	$P < 0.05$
Zebra finch	24	16	19	26	62	NS

NOTE.—NS, not significant.
[a] Sites that were different in all three sequences.

> *Mb* > *GbE*. Under the model in which ω (=d_N/d_S) was allowed to vary among the three paralogous gene lineages, *Cygb* was characterized by the lowest ω value, whereas *GbE* and *Mb* were very similar (table 3). However, forcing all three paralogs *Cygb*, *GbE*, and *Mb* to have a single ω value did not result in a significant change in the likelihood score (supplementary fig. 6, Supplementary Material online). We therefore conclude that these three genes have been evolving under similar levels of purifying selection. The comparative analysis of neognathe birds and boreoeutherian mammals revealed similar patterns of sequence conservation, and the rank order of ω values for the subset of avian and mammalian globin genes that are 1:1 orthologs (*Ngb*, *Cygb*, *Mb*, α^E, α^D, and α^A) was similar in both groups (table 3).

Table 3
Evolutionary Comparisons for Members of the Globin Gene Superfamily in the Genomes of Neognathe Birds (Top) and Boreoeutherian Mammals (Bottom)

	ω[a]	*P* distance (aa)[b]	ω[c]	*P* distance (aa)[b]
	Neognathae[d]		Chicken versus Zebra Finch	
Cygb	0.05	6.2	8.4	0.03
GbE	0.08	6.1	7.3	0.11
α^E-Globin	0.12	8.5	11.3	0.15
α^D-Globin	0.19	19.1	25.5	0.15
α^A-Globin	0.15	5.9	19.0	0.09
ρ-Globin	0.15	9.5	13.6	0.04
B^H-Globin	0.21	13.2	17.0	0.07
β^A-Globin	0.06	5.9	7.5	0.02
ϵ-Globin[f]	0.11	—	—	—
Mb	0.08	5.6	9.1	0.08
Ngb[f]	0.10	—	11.9	0.08
	Boreoeutheria[e]		Human versus Dog	
Cygb	0.03	4.8	4.2	0.09
α^E-Globin	0.15	20.8	17.0	0.12
α^D-Globin	0.11	12.5	15.6	0.09
α^A-Globin	0.17	17.7	12.0	0.08
ϵ-Globin	0.17	13.7	12.4	0.11
β-Globin	0.27	19.1	13.6	0.16
Mb	0.10	16.0	15.6	0.12
Ngb	0.05	5.9	4.6	0.04

NOTE.—aa, amino acid.
[a] ω (=d_N/d_S) estimates from codeml.
[b] Calculated in MEGA, excluding sites with alignment gaps in a pairwise manner.
[c] ω (=d_N/d_S) estimates from yn00.
[d] Neognathae is the clade that includes chicken, duck, turkey, and zebra finch (Supplementary data file 2, Supplementary Material online).
[e] Boreoeutheria is the clade that unites Euarchontoglires and Laurasiatheria. We included sequences from mega- and microchiropteran bats, cow, dog, guinea pig, human, macaque, marmoset, mouse, pika, and rabbit (Supplementary data file 3, Supplementary Material online).
[f] In cases where only two avian sequences were available for comparison, codeml analyses are not reported.

Evolution of the Vertebrate *Cygb*, *GbE*, and *Mb* Genes

We integrated results of the combined phylogenetic and synteny analysis to reconstruct the evolutionary history of the paralogous *Cygb*, *Mb*, and *GbE* genes during the radiation of gnathostome vertebrates (fig. 7). Patterns of conserved macrosynteny demonstrate that the proto-*Cygb* and proto-*GbE*/*Mb* genes represent paralogous products of a whole-genome duplication event in the stem lineage of vertebrates (Storz et al. 2011). Results of the analyses presented here confirm that the vertebrate *GbE* and *Mb* genes represent the paralogous products of a subsequent tandem gene duplication that occurred in the stem lineage of gnathostome vertebrates (fig. 7). We therefore conclude that the *Cygb*, *GbE*, and *Mb* genes were present in the genome of the gnathostome common ancestor. In the branch leading to teleost fish, *GbE* was probably deleted prior to the fish-specific genome duplication (fig. 7). Based on data from the five teleost fish examined here, it appears that only one paralogous *Mb* gene was retained following the fish-specific genome duplication.

The Avian *GbE* Gene: Paradoxical Patterns of Evolution and Enigmatic Function

Our results indicate that the *GbE* gene has been secondarily lost in at least four major gnathostome lineages: teleost fish, amphibians, mammals, and nonavian reptiles. We cannot rule out the possibility that *GbE* has been retained in the genomes of cartilaginous fish, and despite the apparent absence of *GbE* in a representative squamate reptile (anole lizard), we cannot rule out the possibility that the gene has been retained in other nonavian reptiles. It remains to be seen whether *GbE* is found only in birds or whether it is also present in the genomes of other extant archosauromorph taxa (e.g., crocodilians and tuatara). At face value,

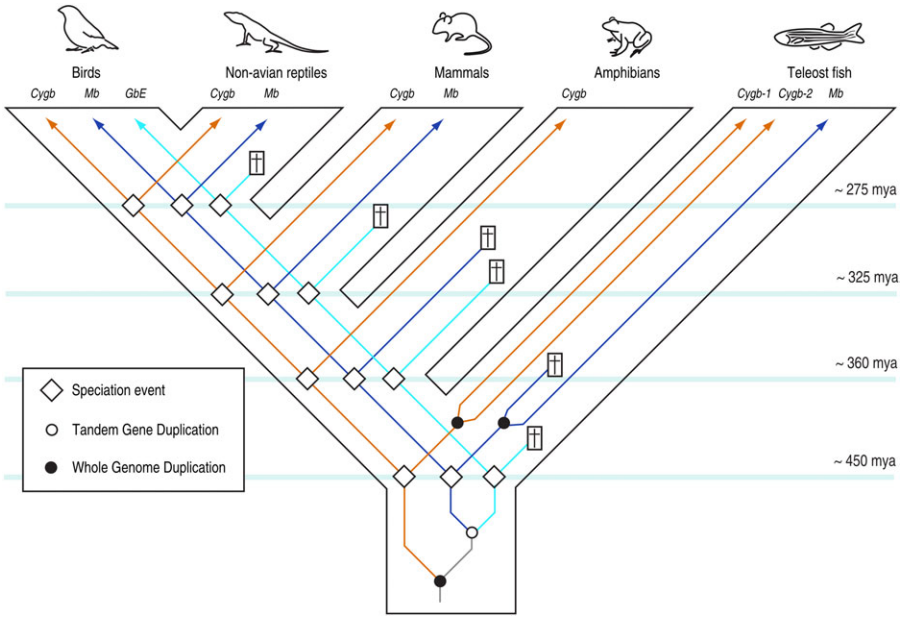

Fig. 7.—A model for the duplicative origins and evolutionary history of the *Cygb*, *GbE*, and *Mb* genes in gnathostome vertebrates. Progenitors of the *Cygb* and *Mb* gene lineages were produced by a whole-genome duplication event in the stem lineage of vertebrates. Subsequently, the *GbE* and *Mb* genes originated via tandem duplication of the proto-*Mb* gene in the common ancestor of gnathostomes (prior to the split between cartilaginous fish and the common ancestor of teleost fish and tetrapods). One or more copies of the *Cygb* gene have been retained in all gnathostome lineages, and a single copy of *Mb* has been retained in all major lineages other than the amphibians, whereas *GbE* has only been retained in the lineage leading to modern birds. The loss of the *GbE* gene in teleost fish probably occurred prior to the fish-specific genome duplication, which may have given rise to the *Cygb-1* and *Cygb-2* paralogs. Estimated divergence dates are taken from Hedges (2009) and Shedlock and Edwards (2009).

the low rate of retention of *GbE* among gnathostome lineages would seem to suggest that the gene is not performing an essential physiological function. Paradoxically, however, relative rate tests revealed that the avian *GbE* gene is characterized by a higher average level of amino acid sequence conservation than the *Mb* gene (table 2), and in orthologous comparisons among all members of the globin gene superfamily in birds, the *GbE* gene was characterized by higher levels of sequence conservation than all genes other than *Ngb*, *Cygb*, and β^A-globin (table 3). We therefore conclude that *GbE* is performing an important physiological function and that function appears to be unique to the avian eye (Kugelstadt et al. 2004).

Other globin proteins such as Ngb and Cygb are also expressed in the retina of birds and other gnathostome vertebrates (Burmester et al. 2002, 2004; Kugelstadt et al. 2004; Fuchs et al. 2005, 2006; Hankeln et al. 2005; Xi et al. 2007; Hankeln and Burmester 2008; Burmester and Hankeln 2009). It has been suggested that these globins may perform an oxygen supply function to sustain mitochondrial respiration in the inner segments of retinal photoreceptors and other aerobically metabolizing cell types (Schmidt et al. 2003, 2005; Bentmann et al. 2005; Hankeln and Burmester 2008; Burmester and Hankeln 2009). For example, globins such as Ngb, Cygb, and GbE that are expressed in the retina could be performing Mb-like respiratory functions by buffering transient fluctuations in cellular oxygen tension or by facilitating intracellular oxygen diffusion to the mitochondria (Sun et al. 2001, 2003). Alternatively (but also in analogy with Mb function), these globins could play a role in the detoxification of reactive oxygen and nitrogen species, they could help maintain nitric oxide (NO) homeostasis, or they could supply oxygen to cytoplasmic enzymes that catalyze oxygen-requiring reactions, such as NO production by NO synthases (Fago et al. 2004; Schmidt et al. 2004; Weber and Fago 2004; Hankeln et al. 2005; Fordel et al. 2007; Hankeln and Burmester 2008; Burmester and Hankeln 2009). Experimental efforts to characterize the structure, reactivity, and cellular localization of GbE will likely provide additional clues to the physiological function of this protein. The observed levels of sequence conservation suggest that such experiments would be worthwhile because whatever GbE is doing, it appears to be something important.

Supplementary Material

Supplementary figures 1–6 and supplementary tables 1–3 are available at *Genome Biology and Evolution online* (http://gbe.oxfordjournals.org/).

Acknowledgments

We thank Z. Cheviron, J. Projecto-Garcia, and three anonymous reviewers for helpful comments and suggestions.

This work was funded by grants to J.F.S. from the NSF (IOS-0949931) and NIH/NHLBI (R01 HL087216 and HL087216-S1) and grants to J.C.O. from the Fondo Nacional de Desarrollo Científico y Tecnológico (11080181), the Programa Bicentenario en Ciencia y Tecnología (PSD89), and the Concurso Estadía Jóvenes Investigadores en el Extranjero from the Universidad Austral de Chile.

Literature Cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.

Awenius C, Hankeln T, Burmester T. 2001. Neuroglobins from the zebrafish *Danio rerio* and the pufferfish *Tetraodon nigroviridis*. Biochem Biophys Res Commun. 287:418–421.

Bentmann A, Schmidt M, Reuss S, Wolfrum U, Hankeln T, Burmester T. 2005. Divergent distribution in vascular and avascular mammalian retinae links neuroglobin to cellular respiration. J Biol Chem. 280:20660–20665.

Bininda-Emonds ORP, et al. 2007. The delayed rise of present-day mammals. Nature. 446:507–512.

Brown JW, Rest JS, García-Moreno J, Sorenson MD, Mindell DP. 2008. Strong mitochondrial DNA support for a cretaceous origin of modern avian lineages. BMC Biol. 6:6.

Burge C, Karlin S. 1997. Prediction of complete gene structures in human genomic DNA. J Mol Biol. 268:78–94.

Burmester T, Ebner B, Weich B, Hankeln T. 2002. Cytoglobin: a novel globin type ubiquitously expressed in vertebrate tissues. Mol Biol Evol. 19:416–421.

Burmester T, et al. 2004. Neuroglobin and cytoglobin: genes, proteins and evolution. IUBMB Life. 56:703–707.

Burmester T, Hankeln T. 2009. What is the function of neuroglobin? J Exp Biol. 212:1423–1428.

Burmester T, Weich B, Reinhardt S, Hankeln T. 2000. A vertebrate globin expressed in the brain. Nature. 407:520–523.

Ebner B, et al. 2010. The globin gene family of the cephalochordate amphioxus: implications for chordate globin evolution. BMC Evol Biol. 10:370.

Ericson PGP, et al. 2006. Diversification of Neoaves: integration of molecular sequence data and fossils. Biol Lett. 2:543–547.

Fago A, Hundahl C, Malte H, Weber RE. 2004. IUBMB Life. 56:689–696.

Fordel E, Thijs L, Moens L, Dewilde S. 2007. Neuroglobin and cytoglobin expression in mice. Evidence for a correlation with reactive oxygen species scavenging. FEBS J. 274:1312–1317.

Fuchs C, Burmester T, Hankeln T. 2006. The amphibian globin gene repertoire as revealed by the *Xenopus* genome. Cytogenet Genome Res. 112:296–306.

Fuchs C, et al. 2004. Zebrafish reveals different and conserved features of vertebrate neuroglobin gene structure, expression pattern, and ligand binding. J Biol Chem. 279:24116–24122.

Fuchs C, Luckhardt A, Gerlach F, Burmester T, Hankeln T. 2005. Duplicated cytoglobin genes in teleost fishes. Biochem Biophys Res Commun. 337:216–223.

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

Hankeln T, Burmester T. 2008. Neuroglobin and cytoglobin. In: Gosh A, editor. The smallest biomolecules: diatomics and their interactions with heme proteins. Amsterdam: Elsevier. pp. 203–218.

Hankeln T, et al. 2005. Neuroglobin and cytoglobin in search of their role in the vertebrate globin family. J Inorg Biochem. 99: 110–119.

Hedges S. 2009. Vertebrates (Vertebrata). In: Kumar S, Hedges S, editors. The timetree of life. New York: Oxford University Press. pp. 309–314.

Hemmingsen E. 1991. Respiratory and cardiovascular adaptation in hemoglobin-free fish: resolved and unresolved problems. In: di Prisco G, Maresca B, Tota B, editors. Biology of antarctic fish. New York: Springer-Verlag. pp. 191–203.

Hoffmann FG, Opazo JC, Storz JF. 2008a. Rapid rates of lineage-specific gene duplication and deletion in the α-globin gene family. Mol Biol Evol. 25:591–602.

Hoffmann FG, Opazo JC, Storz JF. 2008b. New genes originated via multiple recombinational pathways in the β-globin gene family of rodents. Mol Biol Evol. 25:2589–2600.

Hoffmann FG, Opazo JC, Storz JF. 2010. Gene cooption and convergent evolution of oxygen transport hemoglobins in jawed and jawless vertebrates. Proc Natl Acad Sci U S A. 107: 14274–14279.

Hoffmann FG, Storz JF, Gorr TA, Opazo JC. 2010. Lineage-specific patterns of functional diversification in the α- and β-globin gene families of tetrapod vertebrates. Mol Biol Evol. 27:1126–1138.

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

Katoh K, Asimenos G, Toh H. 2009. Multiple alignment of DNA sequences with MAFFT. Methods Mol Biol. 537:39–64.

Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform. 9:286–298.

Kugelstadt D, Haberkamp M, Hankeln T, Burmester T. 2004. Neuroglobin, cytoglobin, and a novel, eye-specific globin from chicken. Biochem Biophys Res Commun. 325:719–725.

Kumar S, Nei M, Dudley J, Tamura K. 2008. Mega: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform. 9:299–306.

Kyte J, Doolittle RF. 1982. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 157:105–132.

Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. Mol Biol Evol. 25:1307–1320.

Li C, Orti G, Zhang G, Lu G, Orti G. 2007. A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. BMC Evol Biol. 7:44.

Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. Genetics. 154:459–473.

Lynch M, O'Hely M, Walsh B, Force A. 2001. The probability of preservation of a newly arisen gene duplicate. Genetics. 159:1789–1804.

Maeda N, Fitch WM. 1982. Isolation and amino acid sequence of a monomeric hemoglobin in heart muscle of the bullfrog, *Rana catesbeiana*. J Biol Chem. 257:2806–2815.

Meyer A, Schartl M. 1999. Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. Curr Opin Cell Biol. 11:699–704.

Meyer A, Zardoya R. 2003. Recent advances in the (molecular) phylogeny of vertebrates. Ann Rev Ecol Evol Syst. 34:311–338.

Miller MA, et al. 2009. The CIPRES Portals. CIPRES. Available from: http://www.phylo.org/sub_sections/portal

Muffato M, Louis A, Poisnel C, Crollius HR. 2010. Genomicus: a database and a browser to study gene synteny in modern and ancestral genomes. Bioinformatics. 26:1119–1121.

Murphy WJ, Pringle TH, Crider TA, Springer MS, Miller W. 2007. Using genomic data to unravel the root of the placental mammal phylogeny. Genome Res. 17:413–421.

Nakatani Y, Takeda H, Kohara Y, Morishita S. 2007. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. Genome Res. 17:1254–1265.
Opazo JC, Hoffmann FG, Storz JF. 2008a. Genomic evidence for independent origins of β-like globin genes in monotremes and therian mammals. Proc Natl Acad USA 105:1590–1595.
Opazo JC, Hoffmann FG, Storz JF. 2008b. Differential loss of embryonic globin genes during the radiation of placental mammals. Proc Natl Acad USA 105:12950–12955.
Patel VS, et al. 2008. Platypus globin genes and flanking loci suggest a new insertional model for β-globin evolution in birds and mammals. BMC Biol. 6:34.
Roesner A, Fuchs C, Hankeln T, Burmester T. 2005. A globin gene of ancient evolutionary origin in lower vertebrates: evidence for two distinct globin families in animals. Mol Biol Evol. 22:12–20.
Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.
Schmidt M, et al. 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, et al. 2003. How does the eye breathe? Evidence for neuroglobin-mediated oxygen supply in the mammalian retina. J Biol Chem. 278:1932–1935.
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.
Shedlock A, Edwards S. 2009. Amniotes. (Amniota). In: Kumar S, Hedges S, editors. The timetree of life. New York: Oxford University Press. pp. 375–379.
Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst Biol. 51:492–508.
Sidell BD, O'Brien KM. 2006. When bad things happen to good fish: the loss of hemoglobin and myoglobin expression in antarctic icefishes. J Exp Biol. 209:1791–1802.
Storz JF, Opazo JC, Hoffmann FG. 2011. Phylogenetic diversification of the globin gene superfamily in chordates. IUBMB Life. 63: 313–322.
Sun Y, Jin K, Mao XO, Zhu Y, Greenberg DA. 2001. Neuroglobin is up-regulated by and protects neurons from hypoxic-ischemic injury. Proc Natl Acad Sci U S A. 98:15306–15311.
Sun Y, et al. 2003. Neuroglobin protects the brain from experimental stroke in vivo. Proc Natl Acad Sci U S A. 100:3497–3500.
Swofford DL, Olsen GJ, Waddell PJ, Hillis DM. 1996. Phylogenetic inference. In: Hillis DM, Moritz C, Mable B, editors. Molecular systematics, 2nd ed. Sunderland, MA: Sinauer. pp. 407–514.
Tajima F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics. 135:599–607.
Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y. 2003. Genome duplication, a trait shared by 22000 species of ray-finned fish. Genome Res. 13:382–390.
Taylor JS, Van de Peer Y, Braasch I, Meyer A. 2001. Comparative genomics provides evidence for an ancient genome duplication event in fish. Philos Trans R Soc Lond B Biol Sci. 356:1661–1679.
Trent JT 3rd, Hargrove MS. 2002. A ubiquitously expressed human hexacoordinate hemoglobin. J Biol Chem. 277:19538–19545.
Vilella AJ, et al. 2009. Ensemblcompara genetrees: complete, duplication-aware phylogenetic trees in vertebrates. Genome Res. 19:327–335.
Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview version 2: a multiple sequence alignment and analysis workbench. Bioinformatics. 25:1189–1191.
Weber RE, Fago A. 2004. Functional adaptation and its molecular basis in vertebrate hemoglobins, neuroglobins and cytoglobins. Respir Physiol Neurobiol. 144:141–159.
Wystub S, et al. 2004. Interspecies comparison of neuroglobin, cytoglobin and myoglobin: sequence evolution and candidate regulatory elements. Cytogenet Genome Res. 105:65–78.
Xi Y, Obara M, Ishida Y, Ikeda S, Yoshizato K. 2007. Gene expression and tissue distribution of cytoglobin and myoglobin in the Amphibia and Reptilia: possible compensation of myoglobin with cytoglobin in skeletal muscle cells of anurans that lack the myoglobin gene. Gene 398:94–102.
Yang Z. 2007. Paml 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.

Associate editor: Ross Hardison