Whole-Genome Duplication and the Functional Diversification of Teleost Fish Hemoglobins

Juan C. Opazo,1 G. Tyler Butts,2 Mariana F. Nery,1 Jay F. Storz,3 and Federico G. Hoffmann*2,4

1Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile
2Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University
3School of Biological Sciences, University of Nebraska
4Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University

*Corresponding author: E-mail: federico.g.hoffmann@gmail.com.
Associate editor: Yoko Satta

Abstract

Subsequent to the two rounds of whole-genome duplication that occurred in the common ancestor of vertebrates, a third genome duplication occurred in the stem lineage of teleost fishes. This teleost-specific genome duplication (TGD) is thought to have provided genetic raw materials for the physiological, morphological, and behavioral diversification of this highly speciose group. The extreme physiological versatility of teleost fish is manifest in their diversity of blood–gas transport traits, which reflects the myriad solutions that have evolved to maintain tissue O2 delivery in the face of changing metabolic demands and environmental O2 availability during different ontogenetic stages. During the course of development, regulatory changes in blood–O2 transport are mediated by the expression of multiple, functionally distinct hemoglobin (Hb) isoforms that meet the particular O2-transport challenges encountered by the developing embryo or fetus (in viviparous or oviparous species) and in free-swimming larvae and adults. The main objective of the present study was to assess the relative contributions of whole-genome duplication, large-scale segmental duplication, and small-scale gene duplication in producing the extraordinary functional diversity of teleost Hbs. To accomplish this, we integrated phylogenetic reconstructions with analyses of conserved synteny to characterize the genomic organization and evolutionary history of the globin gene clusters of teleosts. These results were then integrated with available experimental data on functional properties and developmental patterns of stage-specific gene expression. Our results indicate that multiple α- and β-globin genes were present in the common ancestor of gars (order Lepisoteiformes) and teleosts. The comparative genomic analysis revealed that teleosts possess a dual set of TGD-derived globin gene clusters, each of which has undergone lineage-specific changes in gene content via repeated duplication and deletion events. Phylogenetic reconstructions revealed that paralogous genes convergently evolved similar functional properties in different teleost lineages. Consistent with other recent studies of globin gene family evolution in vertebrates, our results revealed evidence for repeated evolutionary transitions in the developmental regulation of Hb synthesis.

Key words: gene duplication, genome duplication, gene family evolution, convergent evolution.

Introduction

Evidence suggests that two successive rounds of whole-genome duplication that occurred early in vertebrate evolution may have played an important role in the evolution of vertebrate-specific innovations (Holland et al. 1994; Meyer 1998; Meyer and Scharf 1999; Shimeld and Holland 2000; Wada 2001; Hoegg and Meyer 2005; Wada and Makabe 2006; Zhang and Cohn 2008; Van de Peer et al. 2009; Hoffmann, Opazo, and Storz 2012). Roughly 320–400 Ma, a third genome duplication occurred in the stem lineage of teleost fish (infraclass Teleostei) following divergence from nonteleost ray-finned fish (Amores et al. 1998, 2011; Postlethwait et al. 2000; Taylor et al. 2001, 2003; Van de Peer et al. 2003; Hoegg et al. 2004; Jaillon et al. 2004; Meyer and Van de Peer 2005; Kasahara et al. 2007; Sato and Nishida 2010). The teleost-specific genome duplication (TGD) is thought to have provided raw materials for the physiological, morphological, and behavioral diversification of teleost fish, perhaps facilitating the radiation of this speciose group into diverse marine and freshwater environments across the planet. Evidence in support of a causal connection between the TGD and phenotypic innovation is provided by studies of TGD-derived gene duplicates that evolved distinct physiological or developmental functions in various teleost lineages (Meyer and Malága-Trillo 1999; Lister et al. 2001; Mulley et al. 2006; Braasch et al. 2006, 2007; Hashiguchi and Nishida 2007; Hoegg and Meyer 2007; Sato and Nishida 2007; Siegel et al. 2007; Yu et al. 2007; Douard et al. 2008; Braasch, Brunet, et al. 2009; Braasch, Volff, et al. 2009; Sato et al. 2009a, 2009b; Arnegard et al. 2010).

The extreme physiological versatility of teleost fishes is manifest in their diversity of blood–gas transport traits (Wells 2009). This diversity reflects the myriad solutions that have evolved to maintain tissue O2 delivery in the face of changing metabolic demands and environmental O2 availability during different ontogenetic stages. Relative to air-breathing vertebrates, fish generally contend with far greater vicissitudes of environmental O2 availability, which is largely because O2 solubility (and hence, the
availability of dissolved O₂ for respiration) varies as a function of water temperature. During ontogeny, regulatory changes in blood-O₂ transport are mediated by the expression of multiple, functionally distinct hemoglobin (Hb) isoforms that are adapted to the particular O₂-transport challenges encountered by the developing embryo or fetus (in viviparous or oviparous species) and in free-swimming larvae and adults (reviewed by Ingermann 1997; Jensen et al. 1998). As in other vertebrates, the developmental regulation of Hb synthesis in fish involves differential expression of duplicated genes that encode the α- and β-chain subunits of distinct tetrameric α₂β₂ Hb isoforms (Chan et al. 1997; Brownlie et al. 2003; Maruyama, Yasumasu, and Iuchi 2004; Maruyama, Yasumasu, Naruse, et al. 2004; Tiede et al. 2011).

Most teleost fish also coexpress functionally distinct Hb isoforms during posthatching life, and these isoforms can be broadly classified (based on electrophoretic mobility at pH > 8.0) as “anodic” or “cathodic.” The anodic Hbs have relatively low O₂ affinities and a pronounced Bohr effect (decreased Hb-O₂ affinity at low pH), whereas the cathodic Hbs have relatively high O₂ affinities, an enhanced responsiveness to allosteric regulation by organic phosphates, and a reversed Bohr effect (increased Hb-O₂ affinity at low pH) in the absence of organic phosphates (Weber and Jensen 1988; Weber 1990, 2000; Jensen et al. 1998; Weber et al. 2000; Wells 2009). Experimental evidence for some species suggests that regulatory changes in intrerythrocytic Hb isoform composition may play a role in the acclimatization response to environmental hypoxia (e.g., Rutjes et al. 2007), but it has not been possible to formulate any broadly consistent empirical generalizations (Weber and Jensen 1988; Weber, 1990, 2000; Ingermann 1997; Wells 2009). A remarkable feature of nearly all anodic Hb isoforms of teleost fish is the Root effect, an extreme reduction in Hb-O₂ binding capacity at low pH, even when blood O₂ tension remains high. The Root effect is considered a key evolutionary innovation in teleost fish, as it plays a critical role in secreting O₂ into the swim bladder for buoyancy control and in supplying O₂ to the avascular retina (Pelster and Weber 1991; Berenbrink et al. 2005; Berenbrink 2007; Wells 2009).

The alpha- and beta-globin genes of jawed vertebrates (Gnathostomata) represent the product of an ancient gene duplication event that occurred roughly 450–500 Ma in the Ordovician, before the divergence between cartilaginous fish (Chondrichthyes) and the common ancestor of ray-finned fish (Actinopterygii) and lobe-finned fish + tetrapods (Sarcopterygii; Goodman et al. 1987; Storz et al. 2011, 2012; Hoffmann, Opazo, and Storz 2012). Subsequent rounds of duplication and divergence gave rise to diverse repertoires of alpha- and beta-like globin genes that are developmentally regulated in different ways in different vertebrate lineages (Hardison 2003; Hoffmann, Storz, et al. 2010). The ancestral linkage arrangement of the alpha- and beta-globin genes is still retained in at least some cartilaginous fish (Marino et al. 2007), teleosts (Chan et al. 1997; Miyata and Aoki 1997; Gillemans et al. 2003; Pisano et al. 2003), and amphibians (Hentschel et al. 1979; Jeffreys et al. 1980; Kay et al. 1980; Hosbach et al. 1983; Fuchs et al. 2006). In amniote vertebrates, by contrast, the alpha- and beta-globin gene clusters are located on different chromosomes due to transposition of the proto beta-globin gene to a new genomic location sometime after the stem lineage of amniotes split from the line leading to amphibians (Hardison 2008; Patel et al. 2008, 2010; Hoffmann, Opazo, and Storz 2012).

The main objective of the present study was to assess the relative contributions of whole-genome duplication, large-scale segmental duplication, and small-scale gene duplication in producing the extraordinary functional diversity of teleost Hbs. To accomplish this, we integrated phylogenetic reconstructions with analyses of conserved synteny to characterize the genomic organization and evolutionary history of the globin gene clusters of teleost fish. These results were then integrated with available experimental data on functional properties and developmental patterns of stage-specific gene expression. Results of the phylogenetic and comparative genomic analyses revealed repeated evolutionary transitions in stage-specific expression in different teleost lineages. Our analyses also revealed that functionally distinct anodic and cathodic adult Hbs evolved independently in different teleost lineages, providing evidence for convergence in the physiological division of labor between coexpressed Hb isoforms.

Materials and Methods

Data Collection

We used bioinformatic techniques to manually annotate the full complement of globin genes in the genomes of six teleost fish available in release 67 of the ensembl database (fugu, Takifugu rubripes; medaka, Oryzias latipes; green spotted puffer, Tetraodon nigroviridis; tilapia, Oreochromis niloticus; three-spined stickleback, Gasterosteus aculeatus; and zebrafish, Danio rerio). We also annotated the globin genes from a live-bearing teleost (platyfish, Xiphophorus maculatus) and a nonteleost ray-finned fish (spotted gar, Lepisosteus oculatus), both available from the Pre!ensembl database. We compared the ensembl data with previous reports on the genomic organization of the globin gene clusters in fugu (Flint et al. 2001), medaka (Maruyama, Yasumasu, and Iuchi 2004; Maruyama, Yasumasu, Naruse, et al. 2004), and zebrafish (Brownlie et al. 2003). We also included coding sequences from the full complement of globin genes from Atlantic cod (Gadus morhua; Borza et al. 2009, Wetten et al. 2010) and Atlantic salmon (Salmo salar; Quinn et al. 2010). However, the fragmentary state of the cod and salmon genome assemblies precluded a detailed comparative analysis of the globin gene clusters in these two species. Finally, we included additional records from tetrapod vertebrates and cartilaginous fish as outgroup sequences for phylogenetic analyses, and we included genomic contigs from representative tetrapods for the purpose of making synteny comparisons. When possible, the annotated genomic sequences were validated by comparison with the relevant expressed sequence tag (EST) databases.
Assessments of Conserved Synteny
To examine patterns of conserved synteny, we annotated the genes found upstream and downstream of the globin gene clusters of seven teleost species (fugu, medaka, platyfish, green-spotted puffer, stickleback, tilapia, and zebrafish) and one nonteleost ray-finned fish (spotted gar). Initial ortholog predictions were derived from the EnsemblCompara database (Vilella et al. 2009) and were visualized using the program Genomicus (Muffato et al. 2010). In addition, 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 Basic Local Alignment Search Tool (BLAST) (Altschul et al. 1990). Partial sequences for genes of interest (representing pseudogenes or artifacts related to incomplete sequence coverage) were identified and annotated with BLAST. To examine large-scale patterns of sequence conservation, we conducted pairwise comparisons of sequence similarity between globin gene clusters using the Pipmaker and Multipipmaker programs (Schwartz et al. 2000, 2003). To facilitate comparisons, genes have been labeled following the Zebrafish Model Organism Database nomenclature guidelines. Finally, we conducted an assessment of conserved synteny between the globin gene clusters of medaka and the reconstructed protokaryotype of the pre-TGD teleost common ancestor provided by Kasahara et al. (2007) and Nakatani et al. (2007).

Sequence Alignment
Separate alignments of the α- and β-globin coding sequences were based on conceptual translations of nucleotide sequences. Alignments were performed using Muscle v 3.8 (Edgar 2004) and the E-INS-i, G-INS-I, and L-INS-i strategies from Mafft v 6.8 (Katoh et al. 2009). We employed MUMSA (Lassmann and Sonnhammer 2005, 2006) to select the best-scoring multiple alignment, and we then used the selected alignment to estimate phylogenetic relationships. These sequence manipulations were carried out in the Mobyle platform server (Néron et al. 2009) hosted by the Institut Pasteur (http://mobyle.pasteur.fr, last accessed September 2012). All sequence alignments are provided in supplementary data file S1, Supplementary Material online.

Phylogenetic Analyses
We reconstructed separate phylogenies for the α- and β-globin genes using Bayesian and maximum likelihood approaches. We performed maximum likelihood analyses in Treefinder, version March 2011 (Jobb et al. 2004), and we evaluated support for the nodes with 1,000 bootstrap pseudoreplicates. We used the “propose model” tool of Treefinder to select the best-fitting models of amino acid and nucleotide substitution, with an independent model for each codon position in analyses based on nucleotide sequences. Model selection was based on the Akaike information criterion with correction for small sample size. We estimated Bayesian phylogenies in MrBayes v 3.1.2 (Ronquist and Huelsenbeck 2003), running six simultaneous chains for $2 \times 10^7$ generations, sampling every $2.5 \times 10^5$ generations, and using default priors. A given run was considered to have reached convergence once the likelihood scores reached an asymptotic value and the average standard deviation of split frequencies remained < 0.01. We discarded all trees that were sampled before convergence, and we evaluated support for the nodes and parameter estimates from a majority rule consensus of the last 2,500 trees.

Results and Discussion
The comparative genomic analysis revealed that the Hb genes of teleost fish are located in two separate chromosomal regions that are clearly delineated by distinct sets of flanking loci (fig. 1). In contrast, the Hb genes of the nonteleost gar are located in a single chromosomal region. Following Hardison (2008), the teleost globin gene cluster flanked by the mpg and nprl3 genes was labeled the “MN” cluster, and the teleost globin cluster flanked by the lcmt1 and aqp8 genes was labeled the “LA” cluster. In the platyfish assembly, we identified two separate scaffolds containing the MN and LA clusters (fig. 1) and a third scaffold (JH559524) that contained a single, putatively functional β-globin gene. We excluded this latter scaffold from all subsequent analyses because it likely represents an assembly artifact. The MN and LA clusters correspond to the medaka E1 and A1 clusters, respectively, that were described by Maruyama et al. (Maruyama, Yasumasu, and Iuchi 2004; Maruyama, Yasumasu, Naruse, et al. 2004). To facilitate comparison, we report the order of genes in the same orientation as they appear in the zebrafish genome assembly, regardless of how they are found in the ensembl database. Since the MN and LA clusters of most teleosts harbor globin genes in both forward and reverse orientations, we use the terms left and right to describe linear gene order. The individual α- and β-globin genes in the MN cluster were numbered from left to right, such that the functional globin gene in the leftmost position of the MN cluster of zebrafish is labeled MN Hbb1, the next gene to the right is MN Hba1, and so forth, whereas the genes in the LA cluster were numbered from right to left, starting with the gene closest to aqp8 (fig. 2). In the case of cod and salmon globin genes, we retained the labels from the original studies (Borza et al. 2009; Quinn et al. 2010). Sequence sources for the globin gene clusters used in this study are provided in table 1, and the annotations for each cluster are provided in supplementary table S1, Supplementary Material online. To facilitate comparisons with previous studies, we compiled a list of previously used names for each of the annotated globin genes (supplementary table S2, Supplementary Material online).

Genomic Structure of the MN and LA Globin Gene Clusters in Teleosts
Patterns of Conserved Synteny
The genomic context of the teleost globin gene clusters is relatively well conserved, especially in the case of the MN cluster. In all teleost species analyzed, there is perfect conservation for the five genes to the left of the MN cluster: aanat,
Fig. 1. Unscaled depiction of the genomic organization of the MN and LA globin gene clusters from representative teleost fishes, with the human α-globin gene cluster provided as reference. To facilitate comparisons, all clusters are presented in the same orientation as the zebrafish. Genes in the forward orientation are shown on top of the chromosome, whereas genes in the reverse orientation are shown below.

Fig. 2. Genomic structure of the MN and LA globin gene clusters of teleost fish. To facilitate comparisons, all clusters are presented in the same orientation as the zebrafish. Genes in the forward orientation are shown on top of the chromosome, whereas genes in the reverse orientation are shown below. The green-spotted puffer globin genes are assumed to have the same stage-specific expression profiles as their orthologous counterparts in fugu. The Hbb pseudogene in the zebrafish MN cluster is not drawn to scale. Gene labels are color coded based on the timing of their expression. Genes marked with an asterisk were not included in the phylogenetic analyses.
With respect to gene content, the number of globin genes ranged from 3.4 kb in stickleback to 17.2 kb in zebrafish. (fig. 2). From the start codon of the first globin gene to the nonfunctional. Comparison with DNA-derived sequence obtained a 4 bp insertion in the second exon that would render substantial variation in both physical extent and gene content used the cDNA-derived sequence for all further analyses. We assumed that the insertion was identical in sequence. We assumed that the insertion was present in all species that lacked the inactivating 4 bp insertion but were otherwise spotted puffer genome (green-spotted puffer LA

Table 1. Data Sources, Genomic Coordinates and Orientations of the Globin Gene Clusters in Fugu, Green-Spotted Puffer, Gar, Medaka, Platypish, Tilapia, Stickleback, and Zebrafish.

<table>
<thead>
<tr>
<th>Species</th>
<th>Release</th>
<th>Cluster</th>
<th>Location</th>
<th>Orientation</th>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fugu (T. rubripes)</td>
<td>Fugu 4.0</td>
<td>LA</td>
<td>Sc_3</td>
<td>lctm1 → rhbdf1b</td>
<td>2,511,982</td>
<td>2,517,373</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MN</td>
<td>Sc_15</td>
<td>kank → nprl3</td>
<td>417,195</td>
<td>420,598</td>
</tr>
<tr>
<td>Green-spotted puffer (Tet. nigroviridis)</td>
<td>46</td>
<td>LA</td>
<td>Chr 2</td>
<td>rhbdf1b → aap8</td>
<td>5,887,638</td>
<td>5,893,221</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MN</td>
<td>Chr 3</td>
<td>kank → nprl3</td>
<td>12,162,093</td>
<td>12,165,924</td>
</tr>
<tr>
<td>Gar (Lepisosteus oculatus)</td>
<td>LepOcu1</td>
<td>Hb</td>
<td>LG13</td>
<td>nprl3 → luc7l</td>
<td>2,809</td>
<td>54,885</td>
</tr>
<tr>
<td>Medaka (O. latipes)</td>
<td>Medaka 1.0</td>
<td>LA</td>
<td>Chr 19</td>
<td>aap8</td>
<td>1,478,030</td>
<td>1,487,664</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MN</td>
<td>Chr 5</td>
<td>nprl3 → kank</td>
<td>8,378,078</td>
<td>8,412,019</td>
</tr>
<tr>
<td>Platypish (X. maculatus)</td>
<td>Xipmac4.4.2</td>
<td>LA</td>
<td>JHS57783</td>
<td>rhbdf1b → aap8</td>
<td>22,438</td>
<td>37,618</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MN</td>
<td>JHS56906</td>
<td>nprl3 → kank</td>
<td>106,543</td>
<td>141,798</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unassigned</td>
<td>JHS59524</td>
<td></td>
<td>5,235</td>
<td>8,503</td>
</tr>
<tr>
<td>Tilapia (Ore. niloticus)</td>
<td>Orenil1.0</td>
<td>LA</td>
<td>GL831136</td>
<td>rhbdf1b → aap8</td>
<td>111,303</td>
<td>122,995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MN</td>
<td>GL831149</td>
<td>nprl3 → kank</td>
<td>110,462</td>
<td>169,554</td>
</tr>
<tr>
<td>Stickleback (G. aculeatus)</td>
<td>BROADS1</td>
<td>LA</td>
<td>Sc 112</td>
<td>c17orf28 → aap8</td>
<td>339,530</td>
<td>343,463</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MN</td>
<td>Gr XI</td>
<td>kank → nprl3</td>
<td>13,640,461</td>
<td>13,663,356</td>
</tr>
<tr>
<td>Zebrafish (D. rerio)</td>
<td>Zv9</td>
<td>LA</td>
<td>Chr 12</td>
<td>rhbdf1b → aap8</td>
<td>21,688,806</td>
<td>21,705,956</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MN</td>
<td>Chr 3</td>
<td>nprl3 → kank</td>
<td>55,938,147</td>
<td>55,999,373</td>
</tr>
</tbody>
</table>

Note.—In all cases, data were obtained from Ensembl. The start and end points correspond to the most distant edges from the two genes on either end of the cluster.

Hb Gene Repertoires of Teleost Fish

There were several cases where our manual annotations of the globin gene clusters differed from annotations provided in the most recent releases of the various teleost genome assemblies. For example, no MN-linked globin genes were annotated in the most recent release of the fugu genome in the ensembl database. However, BLAST comparisons with an independent record of the fugu MN globin cluster (AY016024) revealed the presence of two unannotated α-globin genes between nprl3 and kank2, as reported by Flint et al. (2001). In addition, the only annotated β-globin gene in the green-spotted puffer genome (green-spotted puffer LA Hbb1) contained a 4 bp insertion in the second exon that would render it nonfunctional. Comparisons with cDNA-derived sequence databases revealed several putatively functional transcripts that lacked the inactivating 4 bp insertion but were otherwise identical in sequence. We assumed that the insertion was either a sequencing or assembly artifact, and we therefore used the cDNA-derived sequence for all further analyses.

The MN and LA clusters of the different species exhibited substantial variation in both physical extent and gene content (fig. 2). From the start codon of the first globin gene to the stop codon of the last globin gene, the MN cluster ranged from 3.4 kb in fugu to 68.5 kb in zebrafish, and the LA cluster ranged from 3.4 kb in stickleback to 17.2 kb in zebrafish. With respect to gene content, the number of globin genes in these clusters ranged from 2 in the MN clusters of fugu and green-spotted puffer and the LA cluster of stickleback, to 13 in the MN clusters of tilapia and zebrafish (not including two genes with partial sequence coverage in the tilapia assembly; fig. 2). Interspecific comparisons revealed a higher rate of globin gene turnover in the MN cluster than in the LA cluster. The MN clusters of fugu and green-spotted puffer possess only two α-globin genes in the reverse orientation, whereas the MN clusters of all other teleosts contain interspersed α- and β-globin genes in both head-to-head and head-to-tail orientations (fig. 2). In the case of stickleback, all of the α-globin genes are found in the reverse orientation, and all of the β-globin genes are in the forward orientation (fig. 2). In all other teleosts, in contrast, multiple α- and β-globin genes are found in both forward and reverse orientations. In all species examined, the LA cluster harbors two tandemly duplicated α-globin genes, and when present, the β-globin genes are sandwiched in between the α-globin genes but in the opposite orientation. The comparative genomic analysis revealed that the β-globin genes of stickleback are only present in the MN cluster, whereas the single β-globin genes of fugu and green-spotted puffer are only present in the LA cluster. Thus, the β-globin genes of stickleback and the two tetraodontid species are not 1:1 orthologs. Furthermore, the set of three globin genes in the LA clusters of fugu and green-spotted puffer appear to have been inverted relative to those of medaka, stickleback, and zebrafish. This inversion hypothesis predicts that the LA Hba1 genes from fugu and green-spotted puffer should be most closely related to the LA Hba2 genes of medaka, platypish, stickleback, tilapia, and zebrafish.

The Globin Gene Cluster in Gar and the Origin of the MN and LA Clusters of Teleosts

In most cases, orthologs of genes flanking the MN and LA globin clusters in teleost fish are located in the vicinity of the α-globin gene cluster in human and chicken, which appears to represent the ancestral location of the proto-Hb gene in jawed vertebrates (Hoffmann, Opazo, and Storz 2012). The 2:1
pattern of conserved synteny between teleost fish and tetrapods suggests that the MN and LA globin clusters of teleost fish derive from the TGD, as suggested by Quinn et al. (2010). This inference is also supported by the presence of duplicate copies of rhdbf1 in teleosts, which are co-orthologs of the single-copy rhdbf1 in tetrapods. Additional bioinformatic searches in the vicinity of the globin gene clusters revealed that most teleosts also possess duplicate copies of shisa9 and mkl2, one on the LA cluster and one on the MN cluster, that are co-orthologous to single-copy genes on the same chromosome as the α-globin gene cluster in human and chicken.

Two additional lines of evidence support the hypothesis that the LA and MN clusters represent paralogous products of the TGD. First, we tested the prediction that the spotted gar (a nonteleost ray-finned fish) would possess a single globin gene cluster, since the gar and teleost lineages diverged before TGD. Consistent with this prediction, our comparative genomic analysis revealed that the spotted gar does indeed possess a single globin gene cluster, ∼52 kb in length, that contains 5 α- and 5 β-globin genes in both forward and reverse orientations (fig. 2). The cluster is flanked by copies of cl6or33, polr3k, mgn1, fox1j, aanat, rhdbf1, mpq, and nprl3 on the left, and by copies of luc7l and itfg3 on the right (fig. 1). Second, we tested the prediction that the LA and MN gene clusters of teleosts descend from the same linkage group in the reconstructed protokaryotype of the pre-TGD teleost ancestor. Consistent with this prediction, an analysis of conserved synteny revealed that the MN and LA clusters of medaka are embedded in paralogous chromosomal segments that trace their duplicative origin to chromosome “e” in the pre-TGD teleost protokaryotype inferred by Kasahara et al. (2007) and Nakatani et al. (2007).

Phylogenetic Relationships among Teleost α- and β-Globin Genes

After characterizing the genomic organization of the globin gene clusters in spotted gar and the seven teleost fish, we performed phylogenetic analyses to reconstruct the duplicative history of the α- and β-globin genes. For this analysis, we added the globin gene repertoires of cod and salmon to those of fugu, green-spotted puffer, medaka, platyfish, stickleback, tilapia, and zebrafish, and we also included sequences from representative tetrapods and cartilaginous fish for comparative purposes. All of the different alignment strategies produced very similar results for the α- and β-globin data sets, and in both cases, we selected the L-INS-i alignment for use in the phylogenetic reconstructions because it had the highest MUMSA score. Before estimating phylogenies, we selected the best-fitting models of amino acid and nucleotide substitution based on the Akaike information criterion with correction for small sample size. In analyses based on nucleotide sequences, we selected an independent model for each codon position. Results of the model estimation procedure can be found in supplementary table S3, Supplementary Material online.

The estimated phylogenies of vertebrate globin sequences suggested that neither α- or β-globin genes of ray-finned fish are monophyletic relative to their tetrapod counterparts (fig. 3A and B). In the case of α-globin genes, a clade of fish sequences that included a subset of genes derived from the teleost LA cluster (LA Hba1 clade 1 + gar Hba3) were placed sister to the chicken α2-globin gene, whereas all other fish α-globins were placed in a second monophyletic group (fig. 3A). In the case of the β-like globin genes, a clade of two gar sequences, including gar Hbb4 and Hbb5, was placed sister to the chicken β-globins. These arrangements suggest that multiple α- and β-globin genes were present in the common ancestor of Actinopterygii + Sarcopterygii.

The phylogeny shown in figure 3A revealed that fish α-globins can be arranged into two distinct clades, defined by the presence of gar Hba2 and gar Hba3, respectively. In turn, teleost α-globins were arranged into five clades that (with the exception of the cod LA Hba2 sequence) reflect their cluster of origin. The discordant position of the cod LA Hba2 sequence probably represents an assembly artifact. Aside from this cod sequence, all α-globin genes derived from the LA cluster were placed into two strongly supported clades: LA Hba1 clade 1 is sister to gar Hba3, and LA Hba clade 2 is embedded in a strongly supported clade that includes all MN α-globin sequences in addition to LA Hba2 from cod and Hba2, Hba4, and Hba5 from spotted gar (fig. 3A). Genealogical relationships within these two clades of LA α-globins are largely congruent with the known organismal relationships, and in both cases, the deepest split separated the zebrafish genes from those of the remaining euteleost taxa. As expected under the cluster-inversion hypothesis, the leftmost α-globin genes in the LA cluster of fugu and green-spotted puffer are most closely related to the rightmost α-globin genes of medaka, platyfish, stickleback, tilapia, and zebrafish, and vice versa (fig. 3A). Relationships among the α-globin genes in the MN cluster are more complex and are not easily reconciled with the organismal phylogeny. The MN-linked genes are organized into three weakly supported clades (fig. 3A). MN Hba1 clade 1 contains salmon and platyfish sequences in addition to two gar sequences, whereas MN Hba clade 2 contains zebrasfish and salmon sequences. MN Hba clade 3 was placed sister to LA Hba1 clade 2 and includes sequences from all teleosts in addition to cod LA Hba2. All species examined possess an α-globin gene repertoire that includes representatives of at least three of the five clades, and zebrafish possesses α-globin genes that are represented in four of the five clades.

In contrast to the α-globin genes, all teleost β-globin genes were placed in a moderately well-supported clade, which was placed sister to a clade of two gar Hbb sequences (Hbb1 and Hbb2). The other two gar Hbb sequences were placed sister to chicken Hbb5 (fig. 3B). The β-globin genes could be arranged into four separate clades, three of which were strongly supported, with sequences from the MN cluster forming a paraphyletic group relative to those from the LA cluster. The β-globins from the LA cluster were placed in a monophyletic group, while those from the MN cluster can be grouped into three separate clades, with the exception of Cod MN Hbb1, which is distantly related to the rest. MN Hbb1 clade 1 contains sequences from medaka, salmon, tilapia, and zebrafish; MN
Hbb clade 2 contains sequences from platyfish, salmon, tilapia, and zebrafish; and MN Hbb clade 3 contains sequences from cod, medaka, platyfish, salmon, stickleback, tilapia, and zebrafish (fig. 3B). Within each of these clades, paralogs from the same species almost invariably formed monophyletic groups, which likely reflects a history of lineage-specific duplication, as with the Hba genes of the MN cluster. This is particularly clear in the case of MN Hbb clade 3, where relationships among the different paralogs are congruent with the known organismal phylogeny after accounting for lineage-specific duplications.

With the exception of β-globin genes from the LA cluster, globin genes of the same subunit type from the MN or LA clusters did not form monophyletic groups. Taken together, the analyses of conserved synteny (figs. 1 and 2) and the phylogenetic reconstructions (fig. 3) indicate that the pre-TGD globin gene cluster of teleost fish contained at least two α-globin genes and 2 β-globin genes. Further, the positions of the gar sequences in the phylogenies of α- and β-like globins indicate that multiple globins of each subunit type were present in the common ancestor of gar and teleosts. If further analyses confirm the paraphyly of ray-finned fish α- and β-globins relative to their tetrapod homologs, it would indicate that multiple α- and β-globin genes were present in the common ancestor of Actinopterygii and Sarcopterygii. As for teleosts, after the TGD but before divergence between zebrafish and the remaining euteleost species, one of the two ancestral β-globin paralogs in the LA cluster was secondarily lost such that the post-TGD globin repertoire was reduced from 8 to 7 genes (fig. 4). Similar lineage-specific patterns of gene turnover have been documented in the α- and β-globin gene clusters of mammals and other vertebrates (Hoffmann et al. 2008a, 2008b; Opazo et al. 2008a, 2008b, 2009; Hoffmann, Storz, et al. 2010). On a deeper evolutionary timescale, lineage-specific duplications and deletions have produced extensive variation in the size and membership composition of the globin gene superfamilies among different vertebrate classes and among different
The phylogenies in figure 3 indicate that all salmon α- and β-globin genes are exclusively found in association with MN-linked globin genes from other species. This reflects the fact that salmonid fish have experienced an additional lineage-specific genome-duplication and that all globin genes were deleted from the duplicated LA clusters and were retained exclusively in the duplicated MN clusters (Quinn et al. 2010). With the exception of fugu and green-spotted puffer, which possess identical globin gene repertoires, all other species in our study show evidence of lineage-specific duplications, which are much more frequent in the MN cluster. In fact, aside from fugu and green-spotted puffer, all other species have expanded the repertoire of α- and β-globin genes via lineage-specific duplications. The most striking contrast is between the MN α-globins from platyfish, tilapia, and zebrafish, and the MN β-globins from stickleback. The stickleback β-globins in the MN cluster derive from a recent set of duplications, whereas the α-globins from the MN clusters of platyfish, tilapia, and zebrafish derive from a combination of recent, lineage-specific duplications of genes deriving from more ancient duplications that likely occurred before the TGD.

Repeated Evolutionary Transitions in Functional Properties and Stage-Specific Expression

In light of evidence that the developmental regulation of Hb synthesis has evolved independently in multiple tetrapod lineages (Hoffmann, Storz, et al. 2010; Storz et al. 2011, 2012), we tested for evidence of a similar phenomenon in teleosts by reconstructing phylogenetic relationships among α- and β-like globin genes that are differentially expressed during development. For the purposes of this analysis, globin genes were classified as “early-expressed” if they are preferentially expressed during embryonic or larval developmental stages, whereas genes were classified as “late-expressed” if they are preferentially expressed in juveniles or adults (supplementary table S4, Supplementary Material online). Since fugu and
green-spotted puffer possess a single β-globin gene, we assumed that this gene is expressed during all ontogenetic stages. For comparative purposes, we included additional teleost globins that are known to be preferentially expressed during embryogenesis in channel catfish (*Ictalurus punctatus*; Chen et al. 2010), rainbow trout (*Oncorhynchus mykiss*; Maruyama et al. 1999), and salmon (Leong et al. 2010). We also analyzed late-expressed α- and β-globin genes whose products are incorporated into tetrameric Hbs with highly distinct functional properties, such as the well-characterized anodic and cathodic Hbs of the European eel (*Anguilla anguilla*; Fago et al. 1995, 1997) and dusky notothen (*Trematomus newnesi*; Mazzarella et al. 1999). Expression data for cod, medaka, and zebrafish were obtained from the literature. The cod sequences were classified following Wetten et al. (2010), the medaka sequences were classified following Maruyama, Yasumasu and Iuchi (2004), and the zebrasfish sequences were classified following Tiedke et al. (2011). For globin genes in fugu, gar, green-spotted puffer, platyfish, salmon, and tilapia, we inferred the timing of expression by identifying matches with sequences in EST databases (supplementary table S4, Supplementary Material online). In the cases of sequences with no matches from the same species as the query sequence or lack of developmental information for the EST matches, the sequences were left as unclassified.

Intriguingly, results of our analyses revealed repeated evolutionary transitions in stage-specific expression during development. In some cases, paralogous genes in different species evolved convergent expression patterns, and in other cases, orthologous genes evolved divergent expression patterns. In the case of the α-like globin genes, LA Hba clades 1 and 2 provide clear examples of probable 1:1 orthologs that evolved differences in stage-specific expression (e.g., the early-expressed zebrafish LA Hba1 and the late-expressed medaka LA Hba1; fig. 3A). In the case of β-like globin genes, LA Hbb clade 1 illustrates a similar pattern of replicated expression divergence (e.g., the early expressed zebrasfish LA Hbb1 and Hbb2 genes are clearly co-orthologous to the adult-expressed medaka LA Hbb1; fig. 3B). These results demonstrate that the developmental timing of globin gene expression is evolutionarily labile.

In the α- and β-globin gene clusters of most amniotes, the linear order of the genes reflects their temporal order of expression during development, with early-expressed genes at the 5’ end of the cluster and late-expressed genes at the 3’ end of the cluster (Hardison 2001). In the globin gene clusters of teleosts, in contrast, gene order is not as strong a predictor of stage-specific expression. In the case of the zebrafish MN cluster, all late-expressed genes are on the left and all early-expressed genes are on the right, whereas in medaka, all genes on the left side are early-expressed and the genes on the right are variable with respect to the developmental timing of expression, and in tilapia, the early- and late-expressed genes are interspersed. Our results indicate that the genes in the LA cluster provide the clearest evidence of lineage-specific changes in gene expression.

Since embryonic/fetal Hbs and adult-expressed Hbs exhibit consistent differences in O₂-affinity and sensitivity to allosteric regulators (Ingermann 1997), convergence in stage-specific expression also likely entailed convergence in functional properties. Similarly, adult α- and β-globin genes that encode the subunits of cathodic Hbs of European eel and dusky notothen are clearly not 1:1 orthologs (fig. 6), indicating that specialized Hbs with similar functional properties evolved independently in different teleost lineages. In fact, the dusky notothen cathodic Hba is closely related to sequences in the
LA cluster, whereas the eel cathodic Hba is closely related to sequences to the MN cluster, suggesting they trace their duplicative origin at least to the TGD. Consistent with other studies of vertebrate globins (Berenbrink et al. 2005; Hoffmann et al. 2010), these results demonstrate that similar expression patterns and functional properties in the Hbs of distinct lineages may sometimes represent products of convergent evolution. Although tandemly duplicated globin genes often evolve in concert due to interparalog gene conversion (Hoffmann et al. 2008a, 2008b; Opazo et al. 2009; Runck et al. 2009; Storz et al. 2011), paralogous genes that are products of genome duplications (also known as “ohnologs”) can escape the homogenizing effects of gene conversion because they are located on different chromosomes. This is one possible reason why paralogous gene copies derived from genome duplications may be more likely to diverge in function than tandem gene duplicates.

**Conclusion**

Results of our combined phylogenetic and comparative genomic analyses indicate that some of the teleost α- and β-like globins are representatives of ancient gene lineages, with
duplicative origins that trace back at least to the common ancestor of gar and teleost fish, and potentially back to the common ancestor of Actinopterygii and Sarcopterygii (superglass Osteichthyes). Such a scenario is consistent with the fact that Hb multiplicity has also been documented in cartilaginous fish (Fyhn and Sullivan 1975; Mumm et al. 1978; Weber et al. 1983; Galderisi et al. 1996; Dafre and Reischl 1997). Our results indicate that the common ancestor of ray-finned fish possessed a fairly diverse globin gene repertoire, and in teleosts, this inherited repertoire was further augmented by the TGD, which produced dual sets of α- and β-like globin genes on two paralogous chromosomes. These TGD-derived gene clusters underwent lineage-specific changes in size and membership composition, and the MN gene cluster underwent an especially high rate of gene turnover. The phylogenetic analyses of teleost globins revealed repeated transitions in stage-specific expression patterns, demonstrating a surprising fluidity in the genetic regulatory control of Hb synthesis during development.

Supplementary Material
Supplementary tables S1–S4, figure S1, and data file S1 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments
The authors thank Andi Kautt, Axel Meyer, David A. Ray, and two anonymous reviewers for critical reviews of this manuscript. M.F.N. holds a CONICYT doctoral fellowship. This work was supported by National Science Foundation (EPS-0903787) grant to F.G.H., National Institutes of Health/NHLBI (R01 HL087216 and HL087216-S1) and the National Science Foundation (IOS-0949931) grant to J.F.S., and Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT 1120032) grant to J.C.O.

References


