doi:10.1111/j.1558-5646.2012.01820.x

SUBFUNCTIONALIZATION OF CYPRINID HYPOXIA-INDUCIBLE FACTORS FOR ROLES IN DEVELOPMENT AND OXYGEN SENSING

Kalle T. Rytkönen,[1,2,3] Arash Akbarzadeh,[1,4] Hamed K. Miandare,[1,5] Hiroyasu Kamei,[6,7] Cunming Duan,[6] Erica H. Leder,[1] Tom A. Williams,[8] and Mikko Nikinmaa[1]

[1]Division of Genetics and Physiology, Department of Biology, University of Turku, FI-20014 Turku, Finland

[2]Current Address: Department of Ecology and Evolutionary Biology, Yale Systems Biology Institute, Yale University, 300 Heffernan Drive, West Haven, Connecticut 06516

[3]E-mail: kalle.rytkonen@yale.edu

[4]Department of Fisheries, Faculty of Agriculture and Natural Resources, University of Hormozgan, 3995 Bandar Abbas, Iran

[5]Department of Fishery, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan 4917638585, Iran

[6]Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, Michigan 48109-1048

[7]Current Address: Departments of Animal Sciences and Applied Biological Chemistry, University of Tokyo, Bunkyo-ku, Tokyo 113-8657, Japan

[8]Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne NE2 4HH, United Kingdom

Received April 3, 2012
Accepted September 12, 2012

Among vertebrates, teleost fishes have evolved the most impressive adaptations to variable oxygen tensions in water (Shoubridge and Hochachka 1980; Nilsson and Randall 2010). Under conditions of oxygen deprivation (hypoxia), major changes in gene expression are mediated by hypoxia-inducible factors (HIF alpha). Here we show that *hif alpha* genes were duplicated in the teleost specific whole-genome duplication. Although one of each paralogous gene pair was lost in most teleosts, both copies were retained in cyprinids. Computational analyses suggest that these duplicates have become subfunctionalized with complementary changes in coding and regulatory sequences within each paralogous gene pair. We tested our predictions with comparisons of *hif alpha* transcription in zebrafish, a cyprinid, and sturgeon, an outgroup that diverged from teleosts before the duplication event. Our experiments revealed distinct transcriptional profiles in the cyprinid duplicates: while one of each paralogous pair maintained the ancestral developmental response, the other was more sensitive to changes in oxygen tension. These results demonstrate the subfunctionalization of cyprinid *hif alpha* paralogs for specialized roles in development and the hypoxic stress response.

KEY WORDS: Development, gene duplication, hypoxia-inducible factor, oxygen, teleost fishes, transcription.

Gene duplication followed by neo- or subfunctionalization is a major mechanism by which genes with new functions and associated phenotypic novelties appear (Ohno 1970; Force et al. 1999; Zhang 2003; Lynch and Katju 2004; Conant and Wolfe 2008; Flagel and Wendel 2009). The evolution of metazoan, and particularly vertebrate genomes is marked by an accumulation of transcription factor duplicates (Degnan et al. 2009; Ravi et al. 2009; Vaquerizas et al. 2009). Studies on the divergence of transcription factor function have concentrated on developmental regulators (Prud'homme et al. 2007; Lynch and Wagner 2008; Crow et al. 2009). Far less is known about the evolution of transcription factors mediating responses to environmental stress. This is

EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

© 2012 The Author(s). *Evolution* © 2012 The Society for the Study of Evolution.

surprising because the pathways responsible for development and environmental stress responses are often intimately connected, as exemplified by responses to temperature and oxygen (Dunwoodie 2009; Akerfelt et al. 2010).

The presence of adequate oxygen levels is one of the most crucial environmental factors for animal life. All animal life requires molecular oxygen as the terminal electron acceptor in aerobic energy production, and consequently animals have evolved sophisticated cellular mechanisms to monitor and respond to fluctuations in oxygen availability. Under reduced oxygen supply (hypoxia), major changes in gene expression are mediated by hypoxia-inducible factors (HIF alpha) (Kaelin and Ratcliffe 2008; Lendahl et al. 2009; Semenza 2012). In addition to cellular homeostasis, HIFs are central in embryonic development and organogenesis (Dunwoodie 2009) and have been shown to be under positive selection in some human populations native to high altitudes (Beall et al. 2010; Yi et al. 2010). In mammals, HIF alpha function is usually regulated post-translationally on specific proline residues in an oxygen-dependent degradation (ODD) domain (Chan et al. 2005) by prolyl hydroxylase domain (PHD) enzymes. Additionally, transcriptional control may be central for the regulation of HIFs, especially in fishes (Law et al. 2006; Rahman and Thomas 2007). HIFs exert transcriptional control on gene expression in a range of processes including erythropoiesis, angiogenesis, glucose and iron transport, glycolysis, and cell-cycle control (Wenger et al. 2005; Kaelin and Ratcliffe 2008). All vertebrates have at least three HIF alphas (Rytkonen et al. 2011), HIF-1 alpha, which is the main regulator of the response to acute (short-term) hypoxia; HIF-2 alpha, which is more active in longer-term responses, and HIF-3 alpha, which inhibits the activities of the others (Lendahl et al. 2009).

Oxygen availability has played a particularly important role in shaping the physiological evolution of aquatic animals and the greatest fluctuations of oxygen levels are encountered in fresh water (Nikinmaa 2002). Among vertebrates, water-breathing teleost fishes have evolved a wide array of adaptations to variable oxygen tensions (Shoubridge and Hochachka 1980; Nilsson and Randall 2010). These adaptations are most advanced in the cyprinids (Cyprinidae; minnows, carps, zebrafish, etc.), which comprise species that tolerate hypoxia and even anoxia better than any other vertebrate (Shoubridge and Hochachka 1980; Nilsson and Randall 2010), and are the most species-rich family of freshwater fishes. *Hif alpha* genes have been studied in teleosts, including zebrafish (Kajimura et al. 2006; Rojas et al. 2007; Kopp et al. 2011) and other cyprinids (Law et al. 2006; Rissanen et al. 2006; Shen et al. 2010; Chen et al. 2012) but the evolutionary and functional relationships between these genes are poorly understood.

This study identifies teleost-specific *hif alpha* (hereafter *hif alpha* is usually referred to as *hif*) gene duplications and characterizes the molecular evolution of these gene paralogs. We used computational and experimental approaches to show that these genes were exclusively retained in cyprinids including zebrafish where they have become subfunctionalized for roles in development and the response to hypoxia. Our results have implications for the use of zebrafish as a model species for studying the HIF system and more generally for how teleost HIF paralogs may serve as an evolutionary model of physiologically important functional divergence in transcription factors.

Materials and Methods

SEQUENCES AND SEQUENCE ANALYSIS

Hif alpha coding and 5′-flanking sequences were collected from Ensembl selecting the longest transcripts. This coding sequence collection was supplemented with BLASTN and TBLASTN queries in NCBI nr/nt and est_others databases (Table S1). The novel asp (*Aspius aspius,* JQ027712, JQ027713) sequences were amplified from tissue cDNAs, using degenerate primers designed from aligned fathead minnow EST and zebrafish sequences, and the novel sturgeon (*Huso huso* JQ027714, *Acipenser persicus* JQ027715) sequences were produced with degenerate primers designed on vertebrates (Table S2). Multiple sequence alignments were built with MUSCLE (Edgar 2004) and the protein alignment of first N-terminal 360 amino acids was filtered with Gblocks (Castresana 2000) and then used in PhyloBayes (Lartillot et al. 2009) with the CAT20 model (Quang et al. 2008) to obtain the phylogeny. Additionally, RAxML (Stamatakis et al. 2008) with the JTT+G model was run to obtain independent support values for the phylogeny. Shared genomic gene order of the *hif alpha* flanking genes was evaluated in human, zebrafish, three-spined stickleback, and tetraodon by collecting data from the Ensembl genome browser.

We employed maximum-likelihood estimates of the ratio of nonsynonymous to synonymous substitutions (dN/dS $= w$) to model the selection pressures on selected teleost clades or branches of the whole *hif* phylogeny. This was carried out using codeml in PAML 4 (Yang 2007). Nested likelihood ratio tests (LRTs) were performed for clade-specific tests of selection pressures: $[w(0) \neq w(\text{hif-1A}) = w(\text{hif-1B})]$ versus $[w(0) \neq w(\text{hif-1A}) \neq w(\text{hif-1B})]$ and $[w(0) \neq w(\text{hif-2A}) = w(\text{hif-2B})]$ versus $[w(0) \neq w(\text{hif-2A}) \neq w(\text{hif-2B})]$. To exclude insertion/deletion fragments present in the whole coding sequence (CDS) alignment we ran the model testing with a dataset consisting of the best aligning portion of the genes corresponding to the basic Helix Loop Helix Per Arnt Sim (bHLHPAS) domain, codons 1–360. To obtain branch-specific dN/dS ratios for Figure 1B, we set the selected branches to have free ratios. For this, the analysis was run separately for bHlHPAS domain and shorter parts of ODD domain: 60 codons of N-terminal proline core (NODD) and

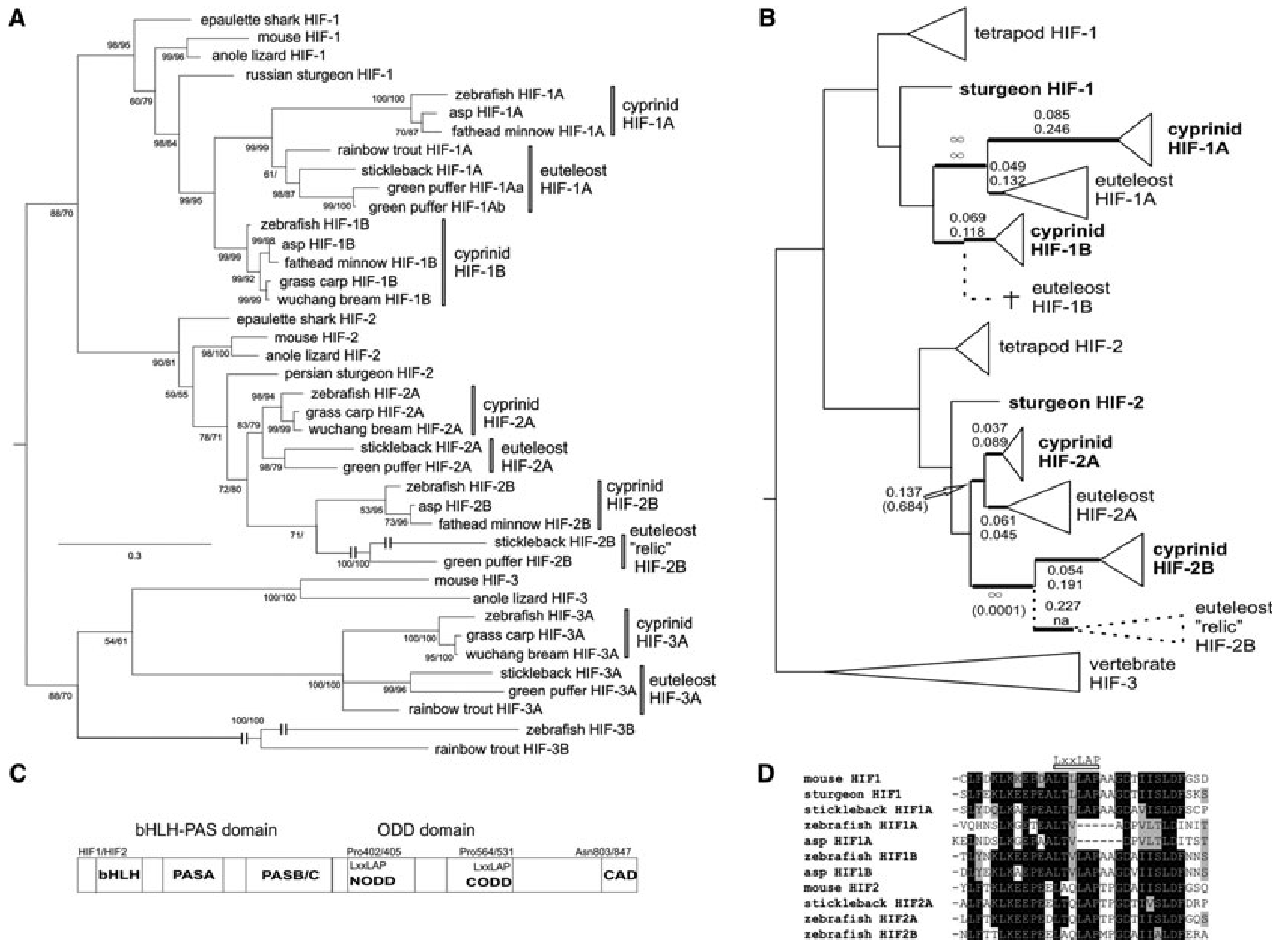

Figure 1. **(A) Bayesian phylogeny of selected HIF alpha genes using PhyloBayes with CAT20 model. The Bayesian posterior probability values and maximum likelihood bootstrap values (RaxML with JTT + G) are displayed beside the branches. The cuts in euteleost partial "relic" HIF-2B and in HIF-3B branches indicate that the long branches were scaled to fit the figure. The unexpected position of teleost HIF-3B is most likely "long branch" artifact caused by extremely high divergence of these gene paralogs. (B) Simplified schematic phylogeny of HIF alpha genes shows that teleost HIF-1A, HIF-1B, HIF-2A, HIF-2B originated from early teleost duplications and that full coding sequences predicting functional HIF-1B and HIF-2B were only retained in a few teleost lineages including cyprinids. The internal branch lengths are from the Bayesian phylogeny above. Free branch-specific nonsynonymous/synonymous (dN/dS) rate ratios obtained from codeml (PAML) are shown, above is the estimate for bHLH-PAS domain and below for the concatenate of NODD and CODD domain cores. Value of ∞ reflects no synonymous substitutions and brackets indicate abnormal dS that may affect dN/dS estimates. The names of the species/gene lineages that were selected for transcriptional studies are in bold. The dashed line indicates the partial "relic" euteleost HIF-2Bs where the CDS contains only the very divergent bHLH-PAS domain and no ODD domain. (C) Schematic overview of the HIF1/2 protein. Basic Helix-Loop-Helix (bHLH) and Per-ARNT-Sim (PASA, PASB/C) or bHLH-PAS-domain is involved in DNA binding and dimerization of the protein. N-terminal oxygen-dependent degradation (NODD) domain and C-terminal ODD (CODD) domain are involved in PHD-dependent degradation. Both of these have the conserved LXXLAP motif. C-terminal transactivation domain (CAD) confers protein–protein interactions to other transcriptional activators. The human positions of crucial hydroxylation sites are depicted. (D) In cyprinid HIF-1A proteins (zebrafish, asp) there is a deletion in the conserved NODD hydroxylation motif (LxxLAP to Lxx—) that may decrease the oxygen sensitivity of the protein.**

60 codons of C-terminal proline core (CODD), corresponding to zebrafish *hif-1A* codons 382–442 and 506—566, respectively.

To obtain independent evolutionary rate estimates we calculated amino acid substitution rates (substitutions/site/year) separately for HIF1 and HIF2 using epaulette shark HIF1 and HIF2 sequences as outgroups. A tetrapod (mouse) was also included in the analysis. Rates were calculated for whole CDS, bHlHPAS domain (1-360 aa) and ODD domains (HIF1 373–596, HIF2 371–559 [mouse]) using divergence time 520 MYA (Blair and Hedges 2005) for the shark—teleost/tetrapod split. For the whole CDS

of zebrafish HIF1A/B and HIF2A/B paralogs a relative rate test (Tajima 1993) implemented in MEGA4 (Tamura et al. 2007) was conducted using shark/sturgeon outgroups.

To predict putative conserved paralog-specific transcription factor binding sites (TFBSs) in proximal promoter (−500 bp), we used “Common TFs” option of MatInspector 8.3 (www.genomatix.de) to gather all the zebrafish A paralog TFBS that are shared with other teleosts (stickleback, fugu, and green puffer) but not with zebrafish B paralogs—and vice versa for B paralogs. The collections of predicted putative conserved TFBS (Table S3) were subtracted from the sister paralog collections to obtain the predicted putative conserved paralog-specific TFBS (Table S4). Additionally, HIF alpha promoters (−500/−1500 bp of transcription start site [TSS]) of selected species were searched for hypoxia response elements (HREs) with MatInspector 8.3.

ANIMAL EXPERIMENTS AND SAMPLING

Zebrafish experiments, sampling, and cDNA production have been described previously (Kamei et al. 2008). Briefly, fertilized eggs were raised in an embryo medium at 28.5°C and staged. Four whole adult zebrafish were exposed to hypoxia for 6 h (oxygen content of 1.8 mg/L versus normoxic oxygen level of 6.2 ∼ 7.1 mg/L), and normoxic controls were included. The staged sturgeon (Beluga, *H. huso*) developmental series was collected in a hatchery and due to its conservation status only surplus material was consumed in this study. Samples were processed as described previously (Akbarzadeh et al. 2011). The zebrafish (z) and sturgeon (s) developmental stages were matched with the following time points: zygote (z: 0–2$\frac{1}{4}$ hpf), blastula (z: 2$\frac{1}{4}$–5$\frac{1}{4}$ hpf), 50% epiboly (z: 5$\frac{1}{4}$–6 hpf), 75–100% epiboly (z: 8 hpf, s: 2 days), 3–6 somite (z: 11–12 hpf), 10–14 somite (z: 14–16 hpf), 21–26 somite (z: 19–22 hpf, s: 4 days), prim-5 (z: 24 hpf), prim-25 (z: 1.5 days, s: 5 days), hatching (z: 2 days, s: 6 days), Protruding mouth (z: 3 days, s: 9 days), morphogenesis (z: 4 days, s: 12 days), morphogenesis completion (z: 5 days, s: 16 days), exogenous feeding (z: 1 week, s: 3 weeks). Additionally, tissue samples were collected from asp (*A. aspius*) and Persian sturgeon (*A. persicus*) for sequencing purposes.

TRANSCRIPTIONAL STUDIES

Zebrafish quantitative PCR (qPCR) primers (Table S2) were designed to span the longest intron using the Roche universal probe system (www.roche-applied-science.com). For the *H. huso hif-2,* the sequence from closely related *A. persicus* was used in primer design. All amplicons were verified by sequencing directly from the qPCR plate after the runs. Reactions were run on 7900 HT Fast Real-Time PCR machine (Applied Biosystems). For zebrafish Kappa Probe Fast qPCR Master Mix (Kappa Biosystems) and for sturgeon Fermentas Maxima SYBR Green qPCR Master Mix (2×) (Fermentas) were used with standard cycling protocol and primers/probes [100 nM]. Conversion of *Ct* values to relative quantities was done using standard curves in ABI 7900HT Version 2.3 SDS (Applied Biosystems). We tested the suitability of four commonly used reference genes (S18 rRNA, eEF1a, beta-acting, RPL6) and as the transcription of all of these was significantly affected by hypoxia, the mRNA expression values were normalized to total RNA. Relative zebrafish mRNA expressions at prim-5 were normalized to 1 for the figures. The zebrafish and sturgeon developmental stages were matched as described above and relative expressions were superimposed by normalizing the highest values from each species. SigmaPlot 11 and SPSS 11 were used for figures and analysis of variance. The *hif-3* developmental data is not shown because zebrafish *hif-3B* transcription was below a reliable detection limit until stage prim-25.

Results and Discussion

LOSS AND RETENTION OF TELEOST HIF ALPHA PARALOGS FOLLOWING WHOLE-GENOME DUPLICATION

We supplemented the available EST and genome data with newly sequenced sturgeon and cyprinid *hif* homologs and built an updated Bayesian molecular phylogeny depicting HIF alpha evolution in teleosts (Fig. 1A and B). Sturgeon is a fish that diverged from the lineage leading to teleosts shortly before the teleost-specific whole-genome duplication (Hoegg et al. 2004). The topology of this tree reveals that all three *hif alpha* genes were duplicated early in the teleost evolution, most parsimoniously in the teleost specific whole-genome duplication, resulting in six *hif alpha* genes: *1A/B, 2A/B,* and *3A/B*. This was followed by the loss of one of each *A/B* paralogous pair in most euteleosts, but not in cyprinids. Additional support for this pattern of retention and loss comes from synteny analysis (Table S5) where the partial coding sequences of *hif-2alphaB* (*hif-2B*) in three-spined stickleback and green puffer paralogs (369 and 308 aa) share more flanking genes with zebrafish *hif-2B* than *hif-2A*. These *hif-2* genes are truncated by an early stop codon for which there is EST support (DT994627, brain; DW645277, larvae), suggesting that no full-length HIF2B is present in these species. Together, these data indicate that the sequenced euteleosts have retained only one functional copy of each pair of *hif-1 alpha* and *hif-2 alpha*, whereas the studied cyprinids possess four intact activating *hif alphas—hif-1A/B* and *hif-2A/B.* With regard to *hif-3 alpha*, we found *hif-3B* paralogs only in zebrafish and rainbow trout (Fig. 1A). In both species *hif-3B* has experienced a very fast rate of evolution, suggesting relaxation of functional constraint or pseudogenization.

The HIF alpha genes of the cyprinid representatives in study thus represent an unusual case in which an ancestral duplication

Table 1. Maximum-likelihood hypothesis testing for clade- and branch-specific selection pressures. Likelihood ratio test were performed to compare if *A* and *B* paralogs within *hif-1* and *hif-2* clades have evolved under differential selection pressures. In both cases the model where both *A* and *B* paralogs were assigned independent dN/dS rate ratio fitted the data better than the model where *A* and *B* paralogs were assigned an equal rate ratio.

Model	w (dN/dS)			ln L	$2\Delta L$
$w(0)$	0.094			−23, 259.44	
$w(0) \neq w(hif\text{-}1A) = w(hif\text{-}1B)$	0.087	0.111		−23, 255.58	
$w(0) \neq w(hif\text{-}1A) \neq w(hif\text{-}1B)$	0.085	0.129	0.091	−23, 250.72	9.72*
$w(0) \neq w(hif\text{-}2A) = w(hif\text{-}2B)$	0.092	0.100		−23, 259.06	
$w(0) \neq w(hif\text{-}2A) \neq w(hif\text{-}2B)$	0.092	0.077	0.143	−23, 251.76	14.60**

*$P < 0.01$; **$P < 0.001$.

has been retained in only one descendent lineage. The cyprinid lineage contains many hypoxia-tolerant species, and we speculated that this ancestral duplication provided the genetic raw material for adaptation to variable oxygen tensions in this lineage. Although the HIF system has been extensively studied in cyprinids (Kajimura et al. 2006; Law et al. 2006; Rissanen et al. 2006; Rojas et al. 2007; Shen et al. 2010; Kopp et al. 2011; Chen et al. 2012) these extra *hif* paralogs have not previously been appreciated because only one of each pair of cyprinid paralogs has been characterized to date (Table S6). We focused our evolutionary analyses on the main activating *hif alphas*, *hif-1A/B* and *hif-2A/B* (Fig. 1B), which show the clearest pattern of differential conservation between cyprinids and other teleosts. These proteins are characterized by two main functional domains: the N-terminal bHLHPAS domain, responsible for the dimerization and DNA binding of HIFs (Scheuermann et al. 2007; Konietzny et al. 2009), and the C-terminal ODD domain that mediates oxygen sensing via interactions with PHD domain enzymes (Chan et al. 2005) (Fig. 1C).

ASYMMETRIC RATES OF EVOLUTION IN TELEOST-SPECIFIC HIF-1 ALPHA AND HIF-2 ALPHA PARALOGS

To test changes in the intensity of natural selection acting on the protein coding sequences of each paralogous pair, we compared the ratio of nonsynonymous to synonymous substitutions (dN/dS) on the clades. In each case, we compared the fit of a model where both sister paralog clades had the same dN/dS ratio to one where they had two different dN/dS ratios using LRTs (Table 1). We found that for both *hif-1A/B* and *hif-2A/B* the model with separate rate ratios for the *A* and *B* clades fitted the data significantly better (*hif-1A/B*, dN/dS: 0.13 and 0.09, $P < 0.01$; *hif-2A/B*, dN/dS: 0.08 and 0.14, $P < 0.001$). Furthermore, a relative rate test (Tajima 1993) on the whole protein coding sequence of zebrafish paralogs using shark/sturgeon outgroups indicated that HIF-1A evolved significantly faster than HIF-1B ($P < 0.001$) and HIF-2B significantly faster than HIF-2A ($P < 0.01$) (see also Table 2). Together, these results imply that the teleost-specific sister paralogs have experienced significantly different selection pressures. Additionally, we obtained free-ratio rate estimates for the selected main internal branches using codeml (PAML) (Fig. 1B), and as expected from the clade-specific tests the *hif-1A* and *2B* branches generally had elevated rate ratio estimates compared to their sister paralogs. In principle, the rate asymmetry may be compatible either with positive selection or relaxed negative selection. However, in the case of cyprinid *hif-1A*, the elevated evolutionary rate (Tables 1 and 2, Fig. 1B) is likely the result of relaxed negative selection: a crucial regulatory motif in the ODD domain of cyprinid *hif-1A,* but not other *hif* paralogs, sustained a 5–6 amino acid deletion (Fig. 1D), suggesting that cyprinid HIF-1A is a less sensitive oxygen regulator than other vertebrate HIF-1 proteins. In addition, amino acid substitution rates are higher within the HIF-1A ODD (1.15 substitutions /site year $\times 10^{-9}$ with shark outgroup) than in HIF-1B (0.77); in contrast, there is no difference in rates between the HIF-2A and 2B ODDs (both 0.98) that are higher than HIF-1B (Table 2), suggesting that HIF-1B is the main acute regulator with intense post-translational surveillance via PHDs. Additionally, when inspecting representative HIF substitution rates in vertebrates (Table 2), uniquely HIF-1B has low substitution rates that are similar to tetrapods (mouse).

The exclusive retention of the ancestral *hif* duplicates in cyprinid representatives, coupled with the rate asymmetry observed in this group, is consistent with an initial neutral accumulation of molecular changes following the ancestral teleost whole-genome duplication that later was coupled with subfunctionalization (Force et al. 1999) or "sub-neo-functionalization" (He and Zhang 2005) of the paralogs, as opposed to rapid neofuntionalization after the ancestral whole-genome duplication followed by stringent negative selection (Ohno 1970). This pattern, in which the functional divergence of transcription factors occurs only in a specific lineage and long after their origin

Table 2. Protein coding and regulatory evolution of HIF alphas. Amino acid substitution rates are shown for zebrafish A and B paralogs and single homologs from stickleback and mouse. All the promoters were analyzed for the number of Hypoxia response elements (HREs) and for the zebrafish promoters the predicted putative conserved paralog specific transcription factor binding sites (TFBS) are shown.

	Subs. rate[1]			Promoter analysis	
	CDS	bHLHPAS	ODD	HREs[2]	Paralog spec. TFBS[3]
HIF1alpha					
Zebrafish HIF1A	0.90	0.59	1.15	(2) 1	EGRF, MZF1
Zebrafish HIF1B	0.65	0.40	0.77	(3) 1	ETSF, NFAT, STAT
Stickleback HIF1	0.78	0.53	0.91	(6) 2	
Mouse HIF1	0.63	0.39	0.72	(5) 2	
HIF2 alpha					
Zebrafish HIF2A	0.78	0.43	0.98	(3) 2	HOXABCD, ZFHX
Zebrafish HIF2B	0.86	0.51	0.98	(8) 6	na
Stickleback HIF2	0.79	0.46	0.91	(11) 2	
Mouse HIF2	0.62	0.34	0.80	(7) 0	

[1]Amino acid substitution rate (substitutions/site year $\times$ 10^{-9}) using molecular estimate of the divergence time for sharks (525 MYA) and epaulette shark sequences as outgroups. bHLHPAS = basic Helix-Loop-Helix – Per-ARNT-Sim; ODD = oxygen-dependent degradation domain.
[2]The number of hypoxia response elements in −1500 bp (in brackets) and −500 bp from TSS.
[3]TFBS specifically conserved in either zebrafish A or B paralog with other teleosts in −500 bp from TSS. EGRF = early growth response factor; MZF1 = myeloid zinc finger factor 1; ETSF = E twenty six factor; NFAT = nuclear factor of activated T cells; STAT = signal transducer and activator of transcription; HOXABCD = hox genes from clusters A,B,C,D; ZFHX = two-handed zinc finger homeodomain transcription factor. See Tables S3 and S4 for details.

in a whole-genome duplication, is emerging as a recurring theme in teleost evolution, with recent examples including the developmental transcription factor HoxA13a in cypriniformes (Crow et al. 2009) and an androgen receptor paralog in percomorphs (Douard et al. 2008).

RECIPROCAL LOSS OF CODING AND REGULATORY ELEMENTS IN CYPRINID HIF ALPHA PARALOGS

Regulation of transcription factor function can evolve either through coding changes in transcription factor genes that alter their activity, specificity or interactions (Lynch and Wagner 2008), or through changes in cis-regulatory elements in noncoding DNA (Prud'homme et al. 2007; Wilson and Odom 2009) that change the mRNA expression of the transcription factor genes and may consequently affect the amount of the factor formed. To complement our analyses of *hif* coding sequences, we investigated the conservation of TFBSs in the proximal promoters (−500 bp) of the zebrafish *hif-1* and *hif-2* paralogs. We included only the "minimal" proximal promoters where the differences between TFBSs are most likely to be functionally critical.

We observed that the proximal promoters of zebrafish *hif-1A* and *2A* have teleost-wide conserved developmentally active motifs, including early growth response and hox motifs that are not present in the *B* paralogs (Tables 2, S3, and S4). Developmental role and transcription during development has been reported in both invertebrates (HIF) (Pocock and Hobert 2008) and vertebrates (HIF-1 and HIF-2) (Rojas et al. 2007; Dunwoodie 2009), so the most parsimonious explanation for the absence of these developmental motifs in B paralogs is that they are retained in *hif-1A* and *2A* and lost in the *B* paralogs. The loss of developmental regulatory motifs in B paralogs together with the relaxed negative selection we observed on the oxygen sensing properties of the cyprinid *hif-1A* coding sequence raise the possibility that cyprinid paralogs may have subfunctionalized for roles during development and in hypoxia response. We hypothesize that the ancestral *hif-1* and *hif-2* were multifunctional genes and that in cyprinids the presence of extra paralogs could have released the constraints imposed by tight oxygen regulation and enabled optimization of the ancestral developmental HIF functions. To test this prediction, we compared the transcript expression of *hif alpha* paralogs during development and in response to hypoxic insult in zebrafish, a tropical cyprinid. For the developmental series, we included sturgeon as an outgroup species to confirm which of the zebrafish *hif-1* and *hif-2* paralogs most resemble the ancestral transcription pattern.

TRANSCRIPTIONAL EVIDENCE FOR SUBFUNCTIONALIZATION OF CYPRINID HIF ALPHA PARALOGS

Our transcriptional results in zebrafish suggest that, consistent with our computational predictions, one member of each paralogous pair (*hif-1* and *hif-2*) is transcriptionally regulated primarily during development, whereas the transcription of the other paralog depends on oxygen status (Fig. 2). The transcription levels

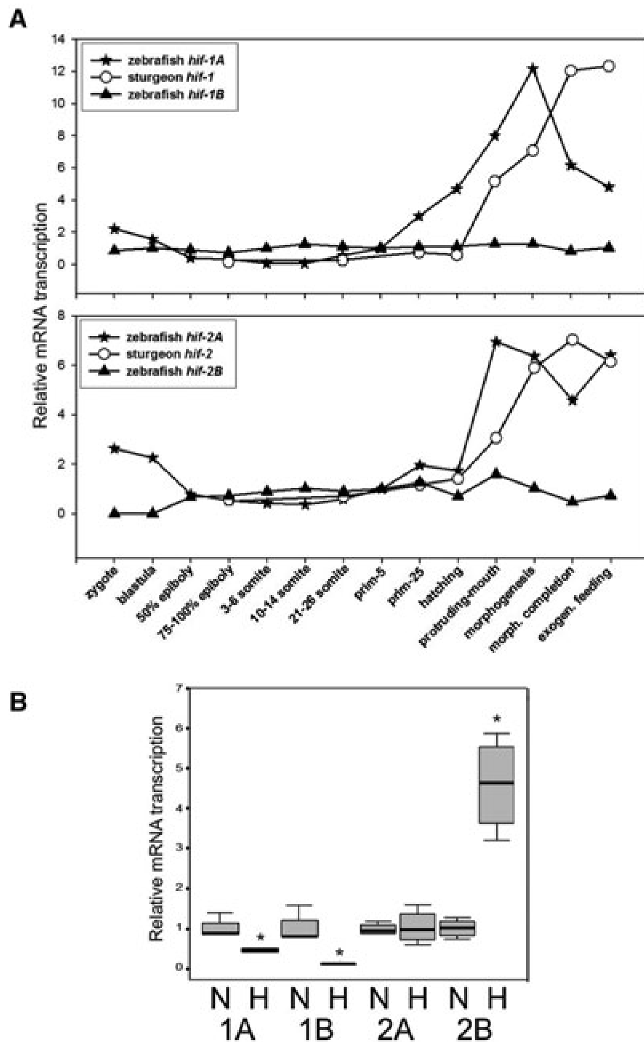

Figure 2. **(A) Relative mRNA expression patterns of *hif alpha* genes during development. Zebrafish paralogs were normalized to prim-5 and sturgeon homologs were superimposed by corresponding developmental stages (see section Materials and Methods) (B) Relative *hif alpha* mRNA expression in normoxic adult zebrafish (N) (normalized to 1) and after 6 h hypoxic (25% air saturation O_2) insult (H), *P < 0.01 ANOVA, (*1A*= *hif-1alpha*A, etc.).**

of *hif-1A* and *hif-2A* both increased approximately 10-fold during development when entering the hatching period, whereas transcription of their paralogs was more affected by hypoxic insult. Hatching is a critical period in fish development and the transcriptional events involved in this process are largely unknown. Combined with the earlier data on the involvement of ARNT2, a dimerization partner of HIF-1 (Hsu et al. 2001) in hatching, our results highlight the possible importance of HIF signaling during this period. Comparison of the zebrafish developmental patterns of *hif-1* and *hif-2* transcription with sturgeon suggested that the developmental upregulation of *hif-1A* and *hif-2A* represents the ancestral regulatory pattern (Fig. 2A).

After hypoxic insult of adult zebrafish, the transcription of *hif-1A* was 40% of the normoxic levels whereas that of *hif-1B* was only 10% (Fig. 2B). The decreased transcription of *hif-1A/B* most likely reflects the generalized downregulation of transcription upon hypoxic shock (Ton et al. 2003; van der Meer et al. 2005) rather than HIF mediated transcriptional feedback loop. Hypoxia did not affect the transcription of *hif-2A*, whereas *hif-2B* transcription was significantly increased ($P < 0.01$). Notably, and consistent with its transcriptional upregulation during hypoxia, the *hif-2B* proximal promoter has more predicted HREs than other HIF genes (Table 2), suggesting also that novel binding sites have contributed to the evolution of teleost HIF system. The transcriptional downregulation of *hif-1* paralogs coupled with upregulation of *hif-2b* after 6 hours of hypoxic insult agrees with previous studies reporting *hif-1* as the acute hypoxia sensitive regulator and *hif-2* as the mediator of long-term hypoxia responses (Holmquist-Mengelbier et al. 2006; Rahman and Thomas 2007). We predict that in fishes the HRE-dependent transcriptional feedback loop of HIFs is directional in such a way that it upregulates transcriptionally only HIF-2, and in the case of zebrafish especially HIF-2B. In the beginning of hypoxic insult HIF-1 (mainly HIF-1B) is the main stabilized (post-translationally affected) HIF paralog and it upregulates the transcription of HIF-2s (mainly HIF-2B) resulting in a HIF-2 dominated response in prolonged hypoxia. It was previously observed that, as in mammals, the HIF-1 protein level is acutely increased in hypoxia in a cyprinid, crucian carp (Rissanen et al. 2006). The time-dependent decrease in HIF-1 protein level after the initial peak is probably due to decreased transcription. In the hypoxia-tolerant cyprinid species the transcriptional decrease of the hypoxia-sensitive HIF-1 gene is pronounced and opposite to that of the transcriptionally hypoxia-responsive HIF-2 gene. This suggests a more sustained elevation of HIF-2 rather than HIF-1 protein level, as increased transcription ensures the maintenance of an elevated protein level that decreases after an initial peak if transcription is decreased, even if protein stability is increased.

CODING VERSUS REGULATORY CHANGES DURING THE EVOLUTION OF HIF ALPHA PARALOGS

Duplicate genes diverge in both coding sequence and expression profile over time (Li et al. 2005), but divergence at these two levels can proceed at different rates and outcomes (Kassahn et al. 2009). In this study, we have shown that both coding and transcriptional changes have contributed the evolution of teleost specific HIF paralogs, but the patterns are different for *hif-1A/B* and *hif-2A/B* sister paralogs. For *hif-1*, the acute hypoxia regulator, both coding and transcriptional changes have apparently contributed to subfunctionalization: in HIF-1A the deletions in the ODD putatively reduce post-translational ODD and transcriptional studies show that *hif-1A*, but not *hif-1B*, mRNA levels vary during

development; in *hif-1B*, in contrast, we observed loss of developmental TFBSs and developmental regulation of transcription, but more intense transcriptional hypoxic regulation. For the "long-term" hypoxia regulator *hif-2* only transcriptional changes are major contributors to subfunctionalization. We detected no indications for functionally relevant changes in the crucial domains of HIF-2A and 2B coding sequences, but the increased number of HREs and loss of developmental TFBSs in the *hif-2B* promoter is consistent with a shift from regulation of development to the hypoxic response. These results, together with developmentally regulated transcription of *hif-2A*, suggest that this paralogous pair has also subfunctionalized.

Conclusions

Our analyses suggest that teleost-specific HIF alpha paralogs have evolved specialized roles in development and in the hypoxia response in zebrafish and other collected cyprinids, whereas most other teleost lineages retain only one copy of each ancestral duplicate. Although the rate asymmetry we detected between the cyprinid paralogs is in principle compatible either with positive selection or relaxed negative selection, our results as a whole imply that the divergence between these paralogs has been driven by loss-of-function mutations consistent with the duplication–degeneration–complementation (DDC) model of subfunctionalization (Force et al. 1999). For *hif-1A/B* we observed that subfunctionalization has proceeded through both coding and transcriptional divergence, whereas for *hif-2A/B* we observed mainly transcriptional divergence between the paralogs. We hypothesize that this subfunctionalization has enabled the optimization of both the developmental and environmental roles of the HIF genes in zebrafish and putatively in other cyprinids, potentially contributing to their radiation in freshwater environments. Considering the importance of HIF alphas in development (Dunwoodie 2009), we anticipate that zebrafish and other cyprinids with subfunctionalized paralogs could be employed as natural models to dissect the developmental versus environmental regulatory roles of HIF signaling, and more generally serve as an evolutionary model of the functional divergence of physiologically important transcription factors following gene duplication.

ACKNOWLEDGMENTS

This work was supported by Academy of Finland and University of Turku. We thank the two anonymous reviewers for their useful comments.

LITERATURE CITED

Akbarzadeh, A., H. Farahmand, F. Mahjoubi, M. A. Nematollahi, P. Leskinen, K. Rytkonen, and M. Nikinmaa. 2011. The transcription of l-gulono-gamma-lactone oxidase, a key enzyme for biosynthesis of ascorbate, during development of Persian sturgeon *Acipenser persicus*. Comp. Biochem. Physiol. B 158:282–288.

Akerfelt, M., R. I. Morimoto, and L. Sistonen. 2010. Heat shock factors: integrators of cell stress, development and lifespan. Nat. Rev. Mol. Cell Biol. 11:545–555.

Beall, C. M., G. L. Cavalleri, L. B. Deng, R. C. Elston, Y. Gao, J. Knight, C. H. Li, J. C. Li, Y. Liang, M. McCormack *et al*. 2010. Natural selection on EPAS1 (HIF2 alpha) associated with low hemoglobin concentration in Tibetan highlanders. Proc. Natl. Acad. Sci. USA 107:11459–11464.

Blair, J. E., and S. B. Hedges. 2005. Molecular phylogeny and divergence times of deuterostome animals. Mol. Biol. Evol. 22:2275–2284.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17:540–552.

Chan, D. A., P. D. Sutphin, S. E. Yen, and A. J. Giaccia. 2005. Coordinate regulation of the oxygen-dependent degradation domains of hypoxia-inducible factor 1 alpha. Mol. Cell. Biol. 25:6415–6426.

Chen, N., L. Chen, J. Zhang, C. Chen, X. Wei, Y. Gul, W. Wang, and H. Wang. 2012. Molecular characterization and expression analysis of three hypoxia-inducible factor alpha subunits, HIF-1α/2α/3α of the hypoxia-sensitive freshwater species, Chinese sucker. Gene 498:81–90.

Conant, G. C., and K. H. Wolfe. 2008. Turning a hobby into a job: how duplicated genes find new functions. Nat. Rev. Genet. 9:938–950.

Crow, K. D., C. T. Amemiya, J. Roth, and G. P. Wagner. 2009. Hypermutability of HoxA13a and functional divergence from its paralog are associated with the origin of a novel developmental feature in zebrafish and related taxa (cypriniformes). Evolution 63:1574–1592.

Degnan, B. M., M. Vervoort, C. Larroux, and G. S. Richards. 2009. Early evolution of metazoan transcription factors. Curr. Opin. Genet. Dev. 19:591–599.

Douard, V., F. Brunet, B. Boussau, I. Ahrens-Fath, V. Vlaeminck-Guillem, B. Haendler, V. Laudet, and Y. Guiguen. 2008. The fate of the duplicated androgen receptor in fishes: a late neofunctionalization event? BMC Evol. Biol. 8:336.

Dunwoodie, S. L. 2009. The role of hypoxia in development of the mammalian embryo. Dev. Cell. 17:755–773.

Edgar, R. C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinf. 5:1–19.

Flagel, L. E., and J. F. Wendel. 2009. Gene duplication and evolutionary novelty in plants. New Phytol. 183:557–564.

Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan, and J. Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545.

He, X. L., and J. Z. Zhang. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics 169:1157–1164.

Hoegg, S., H. Brinkmann, J. S. Taylor, and A. Meyer. 2004. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. J. Mol. Evol. 59:190–203.

Holmquist-Mengelbier, L., E. Fredlund, T. Lofstedt, R. Noguera, S. Navarro, H. Nilsson, A. Pietras, J. Vallon-Christersson, A. Borg, K. Gradin *et al*. 2006. Recruitment of HIF-1 alpha and HIF-2 alpha to common target genes is differentially regulated in neuroblastoma: HIF-2 alpha promotes an aggressive phenotype. Cancer Cell 10:413–423.

Hsu, H. J., W. D. Wang, and C. H. Hu. 2001. Ectopic expression of negative ARNT2 factor disrupts fish development. Biochem. Biophys. Res. Commun. 282:487–492.

Kaelin, W. G., and P. J. Ratcliffe. 2008. Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. Mol. Cell 30:393–402.

Kajimura, S., K. Aida, and C. M. Duan. 2006. Understanding hypoxia-induced gene expression in early development: in vitro and in vivo analysis of hypoxia-inducible factor 1-regulated zebra fish insulin-like growth factor binding protein 1 gene expression. Mol. Cell. Biol. 26:1142–1155.

Kamei, H., L. Lu, S. Jiao, Y. Li, C. Gyrup, L. S. Laursen, C. Oxvig, J. F. Zhou, and C. M. Duan. 2008. Duplication and diversification of the hypoxia-inducible IGFBP-1 gene in zebrafish. PLoS One 3:e3091.

Kassahn, K. S., V. T. Dang, S. J. Wilkins, A. C. Perkins, and M. A. Ragan. 2009. Evolution of gene function and regulatory control after whole-genome duplication: comparative analyses in vertebrates. Genome Res. 19:1404–1418.

Konietzny, R., A. Konig, C. Wotzlaw, A. Bernadini, U. Berchner-Pfannschmidt, and J. Fandrey. 2009. Molecular imaging: into in vivo interaction of HIF-1 alpha and HIF-2 alpha with ARNT. Ann. N.Y. Acad. Sci. 1177:74–81.

Kopp, R., L. Koblitz, M. Egg, and B. Pelster. 2011. HIF signaling and overall gene expression changes during hypoxia and prolonged exercise differ considerably. Physiol. Genom. 43:506–516.

Lartillot, N., T. Lepage, and S. Blanquart. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25:2286–2288.

Law, S. H. W., R. S. S. Wu, P. K. S. Ng, R. M. K. Yu, and R. Y. C. Kong. 2006. Cloning and expression analysis of two distinct HIF-alpha isoforms—gcHIF-1alpha and gcHIF-4alpha—from the hypoxia-tolerant grass carp, *Ctenopharyngodon idellus*. BMC Mol. Biol. 7:15.

Lendahl, U., K. L. Lee, H. Yang, and L. Poellinger. 2009. Generating specificity and diversity in the transcriptional response to hypoxia. Nat. Rev. Genet. 10:821–832.

Li, W. H., J. Yang, and X. Gu. 2005. Expression divergence between duplicate genes. Trends Genet. 21:602–607.

Lynch, M., and V. Katju. 2004. The altered evolutionary trajectories of gene duplicates. Trends Genet. 20:544–549.

Lynch, V. J., and G. P. Wagner. 2008. Resurrecting the role of transcription factor change in developmental evolution. Evolution 62:2131–2154.

Nikinmaa, M. 2002. Oxygen-dependent cellular functions—why fishes and their aquatic environment are a prime choice of study. Comp. Biochem. Physiol. A 133:1–16.

Nilsson, G. E., and D. J. Randall. 2010. Adaptations to hypoxia in fishes. Pp. 131–173 *in* G. E. Nilsson, ed. Respiratory physiology of vertebrates. Cambridge Univ. Press, Cambridge, UK.

Ohno, S. 1970. Evolution by gene duplication. Springer-Verlag, New York.

Pocock, R., and O. Hobert. 2008. Oxygen levels affect axon guidance and neuronal migration in *Caenorhabditis elegans*. Nat. Neurosci. 11:894–900.

Prud'homme, B., N. Gompel, and S. B. Carroll. 2007. Emerging principles of regulatory evolution. Proc. Natl. Acad. Sci. USA 104:8605–8612.

Quang, L. S., O. Gascuel, and N. Lartillot. 2008. Empirical profile mixture models for phylogenetic reconstruction. Bioinformatics 24:2317–2323.

Rahman, M. S., and P. Thomas. 2007. Molecular cloning, characterization and expression of two hypoxia-inducible factor alpha subunits, HIF-1 alpha and HIF-2 alpha, in a hypoxia-tolerant marine teleost, Atlantic croaker (*Micropogonias undulatus*). Gene 396:273–282.

Ravi, V., K. Lam, B. H. Tay, A. Tay, S. Brenner, and B. Venkatesh. 2009. Elephant shark (*Callorhinchus milii*) provides insights into the evolution of Hox gene clusters in gnathostomes. Proc. Natl. Acad. Sci. USA 106:16327–16332.

Rissanen, E., H. K. Tranberg, J. Sollid, G. E. Nilsson, and M. Nikinmaa. 2006. Temperature regulates hypoxia-inducible factor-1 (HIF-1) in a poikilothermic vertebrate, crucian carp (*Carassius carassius*). J. Exp. Biol. 209:994–1003.

Rojas, D. A., D. A. Perez-Munizaga, L. Centanin, M. Antonelli, P. Wappner, M. L. Allende, and A. E. Reyes. 2007. Cloning of hif-1 alpha and hif-2 alpha and mRNA expression pattern during development in zebrafish. Gene Expr. Patterns 7:339–345.

Rytkonen, K. T., T. A. Williams, G. M. Renshaw, C. R. Primmer, and M. Nikinmaa. 2011. Molecular evolution of the metazoan PHD-HIF oxygen-sensing system. Mol. Biol. Evol. 28:1913–1926.

Scheuermann, T. H., J. S. Yang, L. Zhang, K. H. Gardner, and R. K. Bruick. 2007. Hypoxia-inducible factors Per/ARNT/Sim domains: structure and function. Methods Enzymol. 435:3–24.

Semenza, G. L. 2012. Hypoxia-inducible factors in physiology and medicine. Cell 148:399–408.

Shen, R. J., X. Y. Jiang, J. W. Pu, and S. M. Zou. 2010. HIF-1 alpha and -2 alpha genes in a hypoxia-sensitive teleost species *Megalobrama amblycephala*: cDNA cloning, expression and different responses to hypoxia. Comp. Biochem. Physiol. B 157:273–280.

Shoubridge, E. A., and P. W. Hochachka. 1980. Ethanol: novel end product of vertebrate anaerobic metabolism. Science 209:308–309.

Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst. Biol. 57:758–771.

Tajima, F. 1993. Simple method for testing the molecular evolutionary clock hypothesis. Genetics 135:599–607.

Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24:1596–1599.

Ton, C., D. Stamatiou, and C.C. Liew. 2003. Gene expression profile of zebrafish exposed to hypoxia during development. Physiol. Genom. 13:97–106.

van der Meer, D. L., G. E. van den Thillart, F. Witte, M. A. de Bakker, J. Besser, M. K. Richardson, H. P. Spaink, J. T. Leito, and C. P. Bagowski. 2005. Gene expression profiling of the long-term adaptive response to hypoxia in the gills of adult zebrafish. Am. J. Physiol. Regul. Integr. Comp. Physiol. 289:R1512–R1519.

Vaquerizas, J. M., S. K. Kummerfeld, S. A. Teichmann, and N. M. Luscombe. 2009. A census of human transcription factors: function, expression and evolution. Nat. Rev. Genet. 10:252–263.

Wenger, R. H., D. P. Stiehl, and G. Camenisch. 2005. Integration of oxygen signaling at the consensus HRE. Sci. STKE 2005:r12.

Wilson, M. D., and D. T. Odom. 2009. Evolution of transcriptional control in mammals. Curr. Opin. Genet. Dev. 19:579–585.

Yang, Z. H. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591.

Yi, X., Y. Liang, E. Huerta-Sanchez, X. Jin, Z. X. P. Cuo, J. E. Pool, X. Xu, H. Jiang, N. Vinckenbosch, T. S. Korneliussen *et al*. 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. Science 329:75–78.

Zhang, J. Z. 2003. Evolution by gene duplication: an update. Trends Ecol. Evol. 18:292–298.

Associate Editor: T. Near

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Table S1. Accessions of the sequences selected for the phylogeny.
Table S2. Primers used in this study.
Table S3. Summary of the predicted putative transcription factor binding sites (TFBS) in the zebrafish hif alpha promoters (−500 bp from TSS) that have conserved presence in the hif alpha promoters (−500 bp from TSS) of three other fully sequenced teleost species.
Table S4. The predicted putative paralog-specific transcription factor binding sites (TFBS) and hypoxia response elements (HREs) in HIF promoters.
Table S5. Synteny analysis and the flanking genes of vertebrate HIF alphas using data from Ensembl.
Table S6. Zebrafish *HIF alpha* homologs with references to available studies in zebrafish and other cyprinids.