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)

Karen D. Crow,1,2,3 Chris T. Amemiya,4,5 Jutta Roth,3,6 and Günter P. Wagner3

1San Francisco State University, Department of Biology, 1600 Holloway Ave., San Francisco, California 94132
2E-mail: crow@sfsu.edu
3Yale University, Department of Ecology and Evolutionary Biology, 165 Prospect Street, New Haven, Connecticut 06551
4Benaroya Research Institute at Virginia Mason, 1201 Ninth Avenue, Seattle, Washington 98101
5University of Washington, Department of Biology, 24 Kincaid Hall, Seattle, Washington 98195
6Salzburg University, Department of Genetics and General Biology, Hellbrunner Strasse 64, 5020 Salzburg, Austria

Gene duplication is widely regarded as the predominant mechanism by which genes with new functions and associated phenotypic novelties arise (Ohno 1970; Holland and Garcia-Fernández 1996; Meyer and Schartl 1999; Lynch and Katju 2004). Gene and genome duplications are widespread phenomena in eukaryotes (Zhang 2003; Semon and Wolfe 2007a). Early metazoans and chordates have the intriguing tendency to retain duplicate copies of transcription factor gene clusters (Wagner et al. 2003). The duplication, organization, and retention of Hox gene clusters have been the focus of intense research with respect to the significance of genome duplication in the evolution of chordates. Although clearly relevant to the diversification of the chordates,
the functional and evolutionary significance of their retention is not well understood (Lynch and Katju 2004). At the molecular level, duplicate genes provide genetic redundancy that could release one gene copy, or both in part, from purifying selection, allowing evolutionary changes to occur while maintaining the ancestral protein function. In this way, gene duplication may be an important genetic mechanism associated with origin of novel characters (Ohno 1970; Zhang et al. 2002; Zhang 2003) and diversification of species (Zhou et al. 2001; Scannell et al. 2006, 2007; Semon and Wolfe 2007b).

Three whole genome duplications with retention of Hox gene clusters have been documented in the evolutionary history of chordates (Meyer and Schartl 1999), including a genome duplication that occurred shortly before the origin of tetrapods, the most diverse chordate group (Taylor et al. 2003). The majority of tetrapod diversity is sequestered in two main groups—the Osteichthysi and the Perciformes (Nelson 2006). The Osteichthysi is a clade of primarily fresh water fishes, including the zebrafish, *Danio rerio*, and is composed of five groups—the Gonorynchiformes, Characiformes, Siluriformes, Gymnotiformes, and Cypriniformes. Zebrafish belong to the order Cypriniformes—the second most diverse order of ray-finned fishes with approximately 3268 species (Nelson 2006). Because of their experimental tractability, the zebrafish has emerged as a major model organism in the study of vertebrate genetics, development, and evolution. The teleost specific genome duplication resulted in the retention of at least three additional Hox clusters compared to the four Hox clusters observed in other gnathostomes. Among those additional Hox clusters, the co-orthologs of the HoxA cluster, called HoxAa and HoxAb, are the best characterized and provide an ideal opportunity for comparative and experimental investigations (Chiu et al. 2002).

In the zebrafish, there are three pairs of genes from the duplicated HoxA clusters in which both paralogs have been maintained: *HoxA13*, *HoxA11*, and *HoxA9* (although there are detectable pseudogenes for *HoxA10a* and *HoxA2a*). Of these, *HoxA13* is the most divergent and has the highest degree of asymmetry in paralog divergence (Wagner et al. 2005). In this article, we will focus on the molecular and functional divergence of *HoxA13* paralogs in the zebrafish and its relatives. Our goal is to understand when and how these paralogs diverged throughout evolution by mapping evolutionary changes on a phylogeny. Specifically we want to address whether the asymmetric divergence of the paralogs occurred immediately after the teleost-specific duplication. As pointed out by Lynch and Katju (2004) and from published data on the molecular evolution of duplicated genes, it is not clear whether asymmetric divergence and neo-functionalization contribute to the initial retention of paralogs. We show that asymmetric divergence started long after the duplication and probably did not contribute to the initial retention of these paralogs. Second, we want to identify the evolutionary forces leading to asymmetric divergence of *HoxA13* paralogs and present evidence to demonstrate that although synonymous and nonsynonymous substitutions both increase, synonymous substitutions are the predominant class of substitutions contributing the higher rate of evolution in *HoxA13a*. Finally, we test whether divergence of gene function is associated with the asymmetric divergence. We show that in cypriniforms, the faster evolving paralog *HoxA13a*, has many derived, cluster-specific amino acids (highly significant Type II divergence sensu Gu 2006) that are essential for the development of a derived feature in zebrafish - the yolk sac extension (YSE).

**Materials and Methods**

**SEQUENCE DIVERGENCE OF ZEBRAFISH HOXA13 PARALOGS**

**Taxon sampling**

Sequences of *HoxA13* genes were obtained for this study from all major lineages related to the zebrafish in the superorder Osteichthysi, and several basal fish lineages. *HoxA13* sequences from the horn shark *Heterodontus francisci* (Chondrichthyes), the coelacanth *Latimeria menadoensis* (Sarcopterygii), and the basal actinopterygian *Polypterus senegalus* (Polypteriformes) were obtained from public databases (Table 1). New *HoxA13* sequences produced for this study were obtained from single individuals from the following basal actinopterygian and teleost lineages: the gar *Lepisostus platostomous* (Semionotiformes), the bowfin *Amia calva* (Amiiformes), the goldeye *Hiodon alosoides* (Osteoglossomorpha), the eel *Anguilla rostrata* (Elopomorph), and the shad *Dorosoma cepedianum* (Clupeomorph). Within the Osteichthysi, *HoxA13a* and *HoxA13b* paralogs were sequenced from the milkfish *Chanos chanos* (Gonorynchiformes), the knifefish *Apteronotus leptorhynchus* (Gymnotiformes), the catfish *Ictalurus punctatus* (Siluriformes), and four representatives of the Cypriniformes—the zebrafish *Danio rerio*, the goldfish *Carassius auratus*, the algae eater *Gyrinocheilus aymoninii* (Characiformes), and the loach *Misgurnus anguillicaudatus* (Table 1). Sequences from the neon tetra and the Colombian tetra (Characiformes) were produced but ultimately excluded for two reasons: (1) only one paralog sequence was obtained from each taxon and therefore were not as informative for understanding paralog divergence, and (2) because the placement of the Characiformes within the Osteichthysi is contentious, therefore, by excluding the Characiformes we were able to produce a constraint tree that is consistent with currently accepted Osteichthysian relationships.

**DNA extraction and polymerase chain reaction (PCR) amplification**

Muscle or fin tissue was collected and preserved in 95% ethanol. DNA extraction was performed using the DNeasy Tissue Kit
(Qiagen Inc., Valencia, CA) according to the manufacturer’s protocols. Degenerate primers for HoxA13 were designed from conserved regions within exon 1 from sequences of horn shark, coelacanth, bichir, and zebrafish (Table S1). PCR amplification was accomplished using 10 to 100 ng of DNA, 0.2 mM each primer and ReddyMix (ABgene Inc., Rockford, IL) in a final reaction volume of 25 μl. Amplification cycling profiles were generally as follows: 45 sec at 94°C, 45 sec at 52°C, and 1 min at 72°C, ×35 cycles; however, PCR conditions were optimized for each taxon individually. HoxA13 genes were cloned using the pGEM vector system (Promega, Madison, WI). Sequencing was performed with the vector primers T7 or SP6 on an ABI 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA). Several clones of each product were sequenced to assess PCR errors and to increase the probability of detecting duplicated paralogs, resulting in multiple replicate sequences for each gene detected. Sequences for all loci were deposited in GenBank under the following accession numbers: FJ607656-FJ607673 (Table 1).

**Sequence analysis**

Sequences were aligned using the Clustal V algorithm (Higgins et al. 1992), implemented by the software MegAlign (DNASTAR Inc., Madison, WI). Gene genealogies were assessed by neighbor joining (NJ), maximum parsimony (MP), and Bayesian posterior probabilities (BPP) as implemented by the software packages PAUP (version 4.0, Swofford 1998) and MrBayes (version 2.1, Huelsenbeck and Ronquist 2001). The most parsimonious trees were obtained using a heuristic search. Statistical confidence in nodes was evaluated using 2000 bootstrap replicates (Felsenstein 1985; Hedges 1992; Hillis and Bull 1993). For Bayesian analyses, models of evolution were estimated by MrModeltest (Nylander 2002) and statistical confidence in nodes was evaluated by posterior probabilities. Statistical support for nodes is reported as (BPP, 2002) and statistical confidence in nodes was evaluated by posterior probabilities (BPP) as implemented by the software packages PAUP (version 4.0, Swofford 1998) and MrBayes (version 2.1, Huelsenbeck and Ronquist 2001). This analysis controls for phylogeny with a user specified tree in which we judiciously selected a single outgroup lineage (bowfin) according to Robinson-Rechavi et al. (1998). Pairwise relative rate comparisons between HoxA13a and HoxA13b paralogs were estimated by the software package HyPhy (Pond et al. 2005) using the codon model of Goldman and Yang (1994), and the bowfin (Amia calva) as the outgroup. We tested for selection in individual lineages by estimating the nonsynonymous to synonymous substitution rate ratio (dS/dN = w), using codon-based maximum likelihood models of sequence evolution (Goldman and Yang 1994) implemented in the software package PAML (version 3.14, Yang 1997).

**Tests for relative rates of evolution and selection**

Differences in relative rates between whole clades were tested using RRTest (Robinson-Rechavi and Huchon 2000). This analysis controls for phylogeny with a user specified tree in which we judiciously selected a single outgroup lineage (bowfin) according to Robinson-Rechavi et al. (1998). Pairwise relative rate comparisons between HoxA13a and HoxA13b paralogs were estimated by the software package HyPhy (Pond et al. 2005) using the codon model of Goldman and Yang (1994), and the bowfin (Amia calva) as the outgroup. We tested for selection in individual lineages by estimating the nonsynonymous to synonymous substitution rate ratio (dS/dN = w), using codon-based maximum likelihood models of sequence evolution (Goldman and Yang 1994) implemented in the software package PAML (version 3.14, Yang 1997).

**Tests for functional divergence among paralogs**

An amino acid alignment of HoxA13 paralogs was analyzed with the DIVERGE 1.2 software, which implements the methods by Gu (Gu 1999, 2001, 2006). Xun Gu distinguishes two forms of amino acid substitution: degenerate and nondegenerate. The degenerate forms are represented by X and Y, where X is the nondegenerate form and Y is the degenerate form. The nondegenerate forms are represented by A, B, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, and Y. The degenerate forms are represented by *, which is the degenerate form of X.

**Table 1.** Inventory and source and of sequences used for this study with accession numbers. The HoxA13 sequence for the shad, *Dorosoma cepedianum*, is putatively assigned as the HoxA13a* paralog. Cypriniformes shown in bold.

<table>
<thead>
<tr>
<th>Lineage</th>
<th>Order</th>
<th>Code</th>
<th>Scientific name</th>
<th>Common name</th>
<th>HoxA13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chondrichthyes</td>
<td>Hfr</td>
<td>Heterodontus francisci</td>
<td>horn shark</td>
<td>AF479755</td>
<td></td>
</tr>
<tr>
<td>Sarcopterygii</td>
<td>Lme</td>
<td>Latimeria menadoensis</td>
<td>Indonesian coelacanth</td>
<td>AC147788</td>
<td></td>
</tr>
<tr>
<td>Actinoptergii</td>
<td>Pse</td>
<td>Polypterus senegalus</td>
<td>bichir</td>
<td>AC126321</td>
<td></td>
</tr>
<tr>
<td>Semionotiformes</td>
<td>Lpl</td>
<td>Lepisosteus platostomus</td>
<td>shortnose gar</td>
<td>FJ607656</td>
<td></td>
</tr>
<tr>
<td>Amiiformes</td>
<td>Aca</td>
<td>Amia calva</td>
<td>bowfin</td>
<td>FJ607672</td>
<td></td>
</tr>
<tr>
<td>Teleosts</td>
<td></td>
<td></td>
<td></td>
<td>HoxA13a</td>
<td></td>
</tr>
<tr>
<td>Osteoglossomorpha</td>
<td>Osteoglossiformes</td>
<td>Hal</td>
<td>Hiodon alosoides</td>
<td>goldeye</td>
<td>FJ60765</td>
</tr>
<tr>
<td>Elopomorpha</td>
<td>Anguiformes</td>
<td>Aro</td>
<td>Anguilla rostrata</td>
<td>eel</td>
<td>FJ607660</td>
</tr>
<tr>
<td>Clupeomorpha</td>
<td>Clupeiformes</td>
<td>Dce</td>
<td>Dorosoma cepedianum</td>
<td>gizzard shad</td>
<td>FJ607661*</td>
</tr>
<tr>
<td>Ostariophysi</td>
<td>Gonorynchiformes</td>
<td>Chch</td>
<td>Chanos chanos</td>
<td>milkfish</td>
<td>FJ607662</td>
</tr>
<tr>
<td>Ostariophysi</td>
<td>Gymnotiformes</td>
<td>Ale</td>
<td>Apteromonos leptorrhynchus</td>
<td>electric knifefish</td>
<td>FJ607663</td>
</tr>
<tr>
<td>Ostariophysi</td>
<td>Siluriformes</td>
<td>Ipu</td>
<td>Ictalurus punctatus</td>
<td>albino channel catfish</td>
<td>FJ607664</td>
</tr>
<tr>
<td>Ostariophysi</td>
<td>Cypriniformes</td>
<td>Dre</td>
<td>Danio rerio</td>
<td>zebrafish</td>
<td>AC107365</td>
</tr>
<tr>
<td>Ostariophysi</td>
<td>Cypriniformes</td>
<td>Cau</td>
<td>Carassius auratus</td>
<td>goldfish</td>
<td>FJ607665</td>
</tr>
<tr>
<td>Ostariophysi</td>
<td>Cypriniformes</td>
<td>GYR</td>
<td>Gyrinocheilus species</td>
<td>Chinese algae eater</td>
<td>FJ607673</td>
</tr>
<tr>
<td>Ostariophysi</td>
<td>Cypriniformes</td>
<td>Man</td>
<td>Misgurnus anguilicaudatus</td>
<td>dojo loach</td>
<td>FJ607666</td>
</tr>
</tbody>
</table>

*Functional divergence of HoxA13*
acid sequence divergence indicative of functional differentiation: type-I and type-II. Type-I divergence consists of differences in the degree of functional constraints between paralogs, that is, one paralog has a lower rate of amino acid substitutions at certain sites than the other paralog. Type-II divergence consists of amino acid differences that are different and fixed or nearly fixed in both paralogs. The type-II divergence algorithm calculates the site-specific ratio of posterior probabilities of type-II divergence over that of no divergence.

\[ R(F_1 | F_0) = \frac{Pr(F_1 | Data)}{Pr(F_0 | Data)} = r \]

where \( F_1 \) is the hypothesis that the site has experienced type-II divergence and \( F_0 \) that it has not. From which we calculated the posterior probability of type-II divergence (for Fig. S2).

\[ Pr(F_1 | Data) = \frac{r}{1 + r} \]

The probability of false discovery at each site is \( 1 - Pr(F_1 | Data) \) and the estimated total number of true discoveries among all sites was calculated as

\[ E(\# \text{true Type-II sites}) = \sum_{Sites} Pr(F_1 | Data) \]

**FUNCTIONAL PROPERTIES OF ZEBRAFISH HOXA13 PARALOGS**

**Transcriptional activity and artificial recruitment assays**

Full length coding sequences of zebrafish HOXA13a and HOXA13b were obtained by reverse polymerase chain reaction (RT-PCR amplification from zebrafish embryonic mRNA. The parental plasmid for the Hox expression construct was pM2 (Sadowski et al. 1992). PCR constructs were cloned in frame with the GAL4 DNA-binding domain using BamHI and HindIII restriction sites and confirmed by sequencing. Experimental protocols for the artificial recruitment assays closely followed those in Roth et al. (2005). β-galactosidase and luciferase activity were assayed using the β-galactosidase assay system and luciferase assay system, respectively, according to the manufacturer’s instructions (Promega). β-galactosidase activity was measured using an Amer sham (Piscataway, NJ) Ultraspec 2100pro spectrophotometer and luciferase activity was measured using an Analytical Scientific Instruments (El Sobrante, CA) Luminometer 3010, as counts per 10 sec. To normalize differences in transfection efficiencies, luciferase activities were standardized to β-galactosidase levels. Experiments were performed in triplicate.

**Expression patterns of HoxA13 paralogs by real-time PCR during larval development**

Expression patterns of zebrafish HoxA13 paralogs during larval development were assessed using relative quantitative real-time PCR. Broodstock of adult zebrafish (*Danio rerio*) were maintained under standard conditions (Westerfield 2007). Embryos were collected approximately 1 h after onset of the light cycle and presumed to have spawned at the initiation of the light cycle. Fertilized eggs were collected from 4 to 6 spawning chambers with clean marbled bottoms and combined. From this combined pool of embryos, 20–25 individuals were collected at the following time points: 6 h postfertilization (hpf), 12, 18, 24, 30, 36, 42, 48, and 54 hpf.

Total RNA was extracted using the RNeasy Mini Kit from Qiagen Inc. (Valencia, CA). Twelve and a half microliters were used as template for reverse transcription of first-strand cDNA using the Advantage RT-for-PCR Kit (Clontech Laboratories Inc., Mountain View, CA) according to the manufacturer’s protocol. For relative quantitative real-time PCR analysis, approximately 1 μg of cDNA was used in 20 μl reactions with “TaqMan Fast PCR Master Mix” and customized “TaqMan” primer/probe (Applied Biosystems, Foster City, CA). Levels of HoxA13a and HoxA13b expression in larval tissue samples were normalized relative to endogenous levels of β-actin expression in the same sample. Custom primer/probe sets were specifically designed for sequence detection of HoxA13 paralogs and β-actin in zebrafish (Table S1). Thermal cycling conditions for relative quantitative real-time PCRs were as follows: activation, 95°C for 12 sec; denaturation, 95°C for 2 sec, and extension, 62°C for 30 sec; for 60 cycles. All relative quantitative real-time PCRs were carried out on an ABI 7500 Fast Real-Time PCR system with 96-well optical reaction plates (Applied Biosystems), and analyses were performed with the accompanying software. The method for estimating the relative quantitative mRNA levels of the two HoxA13 paralogs is given in the Supporting Information.

**Expression patterns of HoxA13 paralogs via in situ hybridization**

In situ hybridization was performed on zebrafish embryos at six stages of development encompassing the formation of the yolk sac extension including 18, 21, 24, 27, 30, and 36 hpf. Probes were synthesized with the primers A13aProbe and A13bProbe (Table S1) and labeled with digoxigenin. These primers correspond to a 247-bp long region in exon 1 for HoxA13a and a 297-bp region for HoxA13b. The protocol for in situ hybridization was according to (Xu and Wilkinson 1998).

**Knockdown and rescue of zebrafish HoxA13 paralogs**

Translation of HoxA13a and HoxA13b messages was interrupted with morpholinos (MO) designed to target binding sites in the 5’ UTR for each transcript. MO were designed to target the nucleotide sequence 5’ of the first ATG codon to give the most reliable results. These 5’UTRs are paralog specific and well within the ~80 nucleotide position that have been shown to be effective
target regions for MOs to prevent translation. The design of HOXA13aMO1 (Table S1) was based on the published sequence AY303229 described in Amores et al. (2004) and Chiu et al. (2002). This MO targets the first ATG and 22 nucleotides in the 5’ direction. The design of HOXA13bMO1 was based on the cDNA sequence BC090761 and targets the 25 nucleotides immediately 5’ of the first ATG. The corresponding sequence in HOXA13a has 10 mismatches to this sequence. Approximately 1–2 nL of these MOs were injected into the yolk of one- to two-cell stage zebrafish embryos at concentrations of 0.25 mM for the HOXA13a MO and 1.5 mM for HOXA13b MO (Thummel et al. 2004).

For rescue constructs, coding sequences of zebrafish HOXA13a and HOXA13b were cloned into the pCS2+ expression vector from the PM2 phage vector described earlier. The mRNA was transcribed in vitro using the “mMessage mM Machine High Yield Capped RNA Transcription Kit” (Ambion Inc., Austin, TX). After transcription, residual DNA was digested with “TURBO DNase” (Ambion Inc.). The capped, mRNA was purified by LiCl precipitation and resuspended in nuclease free water. Appropriate quantities of mRNA for injection (~10 pg) were determined by titration of serial dilutions injected into 1–2 cell-stage zebrafish embryos and optimized for maximum survivorship with minimal dysmorphic embryos. Microinjections were performed using Eppendorf Femtojet microinjectors and Patchman NP2 micromanipulators and followed the protocol outlined in (Thummel et al. 2004).

All procedures on animal subjects were performed under the guidance and approval of the Institutional Animal Care and Use Committee (IACUC) at Yale University or Benaroya Research Institute at Virginia Mason.

Results

MOLECULAR EVOLUTION OF HOXA13 PARALOGS IN ZEBRAFISH AND RELATED GROUPS

HOXA13 genes were sequenced for 12 taxa including the basal actinopterygians (gar and bowfin), basal teleosts (goldeye, eel, and shad) and representatives of all major lineages within the Ostariophysi. Sequences from the horn shark (Chondrichthyes), and the Indonesian coelacanth (Sarcopterygii) were used as outgroups. We detected one HOXA13 gene in the gar and bowfin, and two paralog sequences in the basal teleost goldeye, and all members of the Ostariophysi considered here, consistent with the teleost specific genome duplication (Hoegg et al. 2004; Crow et al. 2006). Sequences were validated as HOXA13 genes and tentatively identified as orthologous to the known HOXA13 paralogs of zebrafish based on BLAST alignments and gene tree topology. Only one paralog sequence was detected in the eel (HOXA13b), and shad (putatively HOXA13a, but in this case, assessment of orthology was inconsistent depending on taxon sampling and method).

We produced sequences from HOXA13 exon 1 spanning from amino acid 16, the second methionine in the HOXA13 protein, through amino acid 200 for all taxa used in this study. Indels were excluded from further analysis because they had no effect on topology. Gene tree topologies were estimated using the remaining 502 base pairs. Plots of transitions and transversions versus genetic distance indicated that the sequences are not saturated (Fig. S1). The best model of evolution estimated for the complete HOXA13 data set was GTR+I+G according to the Akaike information criteria (AIC) in MrModeltest (Nylander 2002), in which the proportion of invariant sites was estimated to be 0.2272, and the gamma shape parameter (α) was estimated to be 0.9203. The observed transition to transversion ratio was 2.3205.

The NJ tree topology estimated from HOXA13 sequences was nearly identical to the consensus tree estimated using Bayesian methods. The MP analysis estimated two best trees, with a tree score of 1476, that were somewhat less resolved than the NJ and Bayesian consensus tree. The phylogram in Figure 1 indicates posterior probabilities and bootstrap support for MP and NJ at each node.

This topology is consistent with several accepted features of ray-finned fish phylogeny with a few notable exceptions. The predominant features of accepted ray-finned fish topology, with respect to these taxa, include the bichir as the most basal taxon and gar plus bowfin as sister clade to teleosts (Inoue et al. 2003; Kikugawa et al. 2004). Teleosts form a monophyletic clade and their duplicated paralog sequences would separate into well resolved HOXA13a and HOXA13b paralog clades (see constraint tree illustrated in Fig. 2). The shad forms the sister taxon of the Ostariophysi (Inoue et al. 2003). Within the Ostariophysi, the milkfish (Gonorynchiformes) is the most basal and the knifefish (Gymnotiformes) plus catfish (Siluriformes) forms a monophyletic clade sister to the monophyletic Cypriniformes (Saitoh et al. 2003).

The gene tree of HOXA13 is concordant with ray-finned fish phylogeny in that the bichir sequence is in fact the most basal ray-finned fish lineage, and the duplicated HOXA13 genes of the teleosts form a monophyletic clade (100%/57). Within the derived HOXA13a and HOXA13b paralog clades, we recovered monophyletic clades for the Ostariophysi (100/100/100; 100/93/83, respectively) and the Cypriniformes (100/100/100; 100/99/100, respectively), as expected (Fig. 1). However, we did not recover resolved HOXA13a and HOXA13b paralog clades in teleosts as would be expected from a single duplication in the stem lineage of teleosts. In contrast, the HOXA13 paralogs of basal teleosts (gold-eye and eel) are not associated with the more derived paralog groups of the Clupeomorph shad and the Ostariophysi (Fig. 1) but form an unresolved node at the base of the teleost gene lineages. This difference could be due to a number of attributes in the data worth considering, and justifying our use of the constraint tree in Figure 2 for further analysis of sequence evolution.
The gene tree topology could imply that the *HoxA13* genes were duplicated independently in the basal and more derived lineages; however, this topology is not supported by this and multiple other data sets. First, goldeye and eel do not form a monophyletic group according to accepted basal ray-finned fish phylogeny (Inoue et al. 2003). Second, an additional, independent duplication in Ostariophysans is not supported by data sets from other loci that are more clearly resolved. It is more likely that this topology reflects the short period of time between the duplication event and the divergence of the basal teleost lineage for substitutions to accumulate and build up phylogenetic signal at that node, and changes in rate variation across the tree. The duplication history from several *Hox* and non-*Hox* genes support a genome duplication having occurred in the stem lineage of teleosts (Hoegg et al. 2004; de Souza et al. 2005; Crow et al. 2006; Salaneck et al. 2008). In our own study of *HoxA11*, a gene closely linked with *HoxA13*, the node separating the teleost *HoxA11a* and *-b* paralogs was well resolved indicating a single duplication event in the stem lineage of teleosts. The *HoxA11* gene tree was resolved above the gene duplication event because of the accumulation of a number of nonsynonymous substitutions immediately following the duplication event probably by diversifying selection. No synonymous substitutions were detected at the same internode indicating that the time between the gene duplication and the divergence of the basal teleost lineage (Osteoglossomorpha) was short, (between 3.5 and 5 Mio years, Crow et al. 2006). The *HoxA11* paralogs experienced divergent directional selection immediately after the duplication, leading to the accumulation of nonsynonymous differences. The lack of resolution at the base of the *HoxA13* teleost lineages likely reflects a lack of divergent selection on this gene, resulting in the *HoxA13* paralogs of the basal lineages (goldeye and eel) appearing more similar to each other than to the more derived lineages. After the divergence of the shad from the Ostariophysan lineage, the substitution rate increased dramatically, therefore, the *HoxA13a/b* paralogs from the more derived teleost lineages have accumulated enough differences to sort into distinct a and b clades. Further, as a result of the increased substitution rate, the derived lineages are prone to a long-branch attraction artifact, as noted in other cases (Fares et al. 2006). Therefore, we conclude that the most likely true gene tree is the same as for *HoxA11* and other teleost genes (Hoegg et al. 2004; de Souza et al. 2005; Crow et al. 2006; Salaneck et al. 2008), that is, with one duplication event in the stem lineage of the teleost clade. In support of this, we present evidence that the *HoxA13* paralogs did not experience divergent directional selection immediately following the duplication (see below) as observed in *HoxA11*, which, in part, explains the lack of resolution at the base of the teleost *HoxA13* gene lineages.
Figure 2. HOXA13 gene tree constrained to illustrate the currently
accepted phylogenetic relationships of the Ostariophysi (italics,
Saitoh et al. 2003, 2006) and related taxa (Inoue et al. 2003;
Kikugawa et al. 2004), and duplication history of the HOXA13a
(red) and HOXA13b (blue) paralogs. Cypriniforms shown in bold.
Nonsynonymous/synonymous rate ratio (ω) is shown at nodes.
Values of ∞ or zero reflect no synonymous or nonsynonymous
substitutions, respectively. Note that positive selection is not indi-
cated at any internal branches. Effective number of codons (ENC)
indicated by lineage.

OSTARIOPHYSAN HOXA13A EVOLVES FASTER THAN
HOXA13B

The most striking feature of the phylogram illustrated in Figure 1
is the asymmetry in branch lengths between the Ostariophysi
HOXA13a and HOXA13b paralog lineages. To investigate this dif-
fERENCE in rates of evolution between HOXA13 paralogs, we com-
pared relative rates to a basal teleost (Fig. 1) and the more
derived Ostariophysi HOXA13a and HOXA13b paralog clades in
the program RRTree (Robinson-Rechavi and Huchon 2000), using
the bowfin as an outgroup. We found significant differences in the
rate of evolution between the Ostariophysi HOXA13a clade relative
to both the basal teleosts (P = 6.78 × 10−5) and the HOXA13b
paralog clade (P = 0.03096) using the number of nonsynony-
mous transversions per nonsynonymous site (Ba), and correct-
ning for multiple comparisons using the Bonferroni method (n =
3). There was no significant difference in the rate of evolution
between the basal teleosts and the HOXA13b clade after Bonfer-
roni correction (P = 0.14451). Interestingly, the distance-based
method was unable to compute the number of synonymous substi-
tutions per synonymous site (Ks) and number of nonsynonymous
substitutions per nonsynonymous site (Ka). Therefore, the num-
ber of substitutions was too high to compute Ka/Ks ratios, even
though there is no evidence for saturation of transitions (Fig. S1).
In addition to Ba, RRTree estimated the number of synonymous
transversions per fourfold degenerate site (B4). Estimates of B4
were not significantly different between clade comparisons due
to relatively few fourfold degenerate sites. Incidentally, estimates
of Ba still were significantly greater for the HOXA13a relative
to the HOXA13b clade in alternative analyses in which the basal
teleost sequences were excluded (P = 0.025), and when forced to
associate with their respective paralog clades (i.e., user specified
tree reflects Hox cluster duplication before the origin of teleosts,
P = 0.018). Finally, the parameter Ba is the preferred param-
eter in certain circumstances including saturation or differences in
GC content because transversions bias estimates of Ka and Ks
(Robinson-Rechavi and Huchon 2000). However, we found no
evidence for saturation or differences in GC content in HOXA13
paralog sequences (Fig. S1, Table S2).

To further test whether HOXA13a paralogs are evolving at
significantly faster rates than HOXA13b paralogs in zebrafish and
related taxa, we focused on an abbreviated data set including the
bowfin as sister clade to the teleosts, and representatives of all the
major Ostariophysi lineages. This targeted analysis includes se-
quences from the Gonorynchiformes, (milfish, Cheil), the Gym-
notiformes (electric knifefish Ale), the Siluriformes (albino chan-
nel catfish Ipu), and four members of the Cypriniformes (zebrafish
Dre, goldfish Cau, Chinese algae eater Gsp, and the Dojo loach
Man, illustrated in Fig. 3). We found clear evidence for asym-
metric rates of evolution between Ostariophysi HOXA13 paralog
groups. Total branch lengths of the Ostariophysi HOXA13a gene
lineages, from the time of duplication, are consistently longer than
those of HOXA13b lineages. We compared substitution rates (syn-
onymous and nonsynonymous combined) of each HOXA13a par-
alog to all HOXA13b paralog sequences from the Ostariophysi,
using the bowfin, Amia calva, as the outgroup. We found that all
HOXA13a lineages exhibit longer branch lengths than HOXA13b
lineages (with one exception), and 35 of 49 comparisons were sig-
nificant (Table S3). The probability that this consistent asymmetry
would occur by chance is 0.00112.

NO EVIDENCE FOR DIRECTIONAL SELECTION ON THE
FAST EVOLVING HOXA13A

To determine if episodes of directional selection in specific lin-
eages might explain the rate asymmetry between Ostariophysi
HOXA13 paralogs, we used codon-based maximum likelihood
models of sequence evolution (Goldman and Yang 1994; Yang
1997). These methods estimate the nonsynonymous to synony-
mous substitution rate ratios (dN/dS = ω). As an input tree we
used the constraint tree of accepted ray-finned fish phylogeny
(Inoue et al. 2003; Saitoh et al. 2003; Kikugawa et al. 2004), in-
cluding the established origin of zebrafish paralogs resulting from
a whole genome duplication that occurred in the stem lineage of
Figure 3. Estimates of nonsynonymous \(d_N/(NdN)\) and synonymous \(d_S/(SdS)\) substitutions mapped on selected internal lineages of HoxA13 gene genealogy. Rates of synonymous substitutions (shown in bold) are three times higher in HoxA13a paralogs compared to HoxA13b paralogs in zebrafish and relatives. Note that increased rates of evolution affects \(d_N\) as well as \(d_S\).

HoxA13a in Ostariophysans evolves faster than HoxA13b in Cypriniforms (Hoegg et al. 2004; de Souza et al. 2005; Crow et al. 2006; Salaneck et al. 2008). Initial branch lengths were estimated using a one-ratio model (i.e., \(d_N/d_S\) same for all branches) to establish a null model for comparison in more complex analyses. The one ratio model shows that the average nonsynonymous to synonymous rate ratio (\(\omega\)) is 0.0842, indicating strong stabilizing selection averaged over all lineages. Likelihood ratio tests indicate that the two ratio models, in which individual stem lineages have different rate ratios from the rest of the tree, were not significantly better than the one ratio model (i.e., when the stem lineages of Ostariophysans or Cypriniformes were allowed to vary, nor when the lineages within these clades varied). Only the free ratio model, in which all branches vary, was significantly better than the free ratio model \((G = 73.7630, P < 0.001\) based on chi-square approximation). Significant differences in likelihood between the one ratio and free ratio models would indicate heterogeneity in \(d_N/d_S\) ratios among lineages. Heterogeneity may be caused by positive selection or relaxed constraints in purifying selection in some lineages (Yang 1998). Regardless of the cause, heterogeneity of \(d_N/d_S\) ratios indicate selection has not been acting in an equivalent manner among all lineages. A surprising result of this analysis is that there is no evidence that episodes of positive selection on HoxA13a can explain the increased rate of evolution observed in Ostariophysan HoxA13a paralog lineages, although we note that if few sites were under positive selection, such selection would be difficult to detect.

OSTARIOPHYSAN HOXA13A EVOLVES FASTER

Increased rates of evolution are not uncommon in gene lineages derived from gene duplication events (Lynch and Conery 2000; Kondrashov et al. 2002; Conant and Wagner 2003; Wagner et al. 2005), however, the rate asymmetry observed here implies different evolutionary forces acting on HoxA13 paralogs in Ostariophysans and Cypriniforms. An analysis of substitution class and rates using maximum likelihood revealed surprising trends in the pattern of sequence divergence after the gene duplication. Estimates of the number of nonsynonymous (NdN) and synonymous (SdS) substitutions (shown in parentheses in Fig. 3) mapped on the stem lineages between the duplication and the origin of the Ostariophysans and Cypriniforms revealed that both synonymous as well as nonsynonymous substitutions are frequent in the HoxA13a gene lineages than in HoxA13b. But comparing the numbers in detail shows that synonymous substitutions are the predominant cause of increased rate of divergence in HoxA13a. The higher rates of synonymous substitutions explain our inability to detect elevated \(d_N/d_S\) ratios, and both the magnitude of difference between \(d_N\) and \(d_S\) and the observed asymmetry between paralogs was surprising. For example, the rate of synonymous substitutions in the HoxA13a lineages between the gene
duplication (Fig. 3) and the divergence of the milkfish (i.e., origin of the Ostariophysi) is 22 times greater, and between the origin of the Ostariophysi and the origin of Cypriniformes is 34 times greater, than the nonsynonymous substitution rate (Fig. 3). For HoxA13b, the synonymous substitution rates in these lineages are 26 and 14 times greater, respectively. More striking is the observation that overall branch length, the number of SdS, and the rates of synonymous substitutions that occur between the duplication and the origin of cypriniforms are three times greater in the HoxA13a lineage relative to the HoxA13b lineage. There is also an increase in NdN but to a lesser extent. Therefore, asymmetric rates of evolution are primarily explained by differences in rates of SdS between the HoxA13 “a” and “b” paralogs, even though the ages of these paralog lineages are identical due to their common origin at the teleost specific genome duplication.

We considered the following factors that could cause an increase in synonymous substitution rate in HoxA13a compared to b: differences in GC content, differences in mRNA secondary structure, and paralog specific increase in mutation rate. Differences in GC content could lead to differences in synonymous substitution rates (Moriyama and Gojobori 1992; Mouchiroud et al. 1997; Shi et al. 2006). However, we found no difference in GC content big enough to explain the rate difference between Ostariophysan HoxA13 paralog clades (Table S2). If the ancestral mRNAs of HoxA13 had a secondary structure that was lost in HoxA13a but was maintained in HoxA13b, the synonymous sites in HoxA13a would experience fewer selective constraints than in HoxA13b. Again, we found no evidence for secondary structure in either mRNA using a secondary structure prediction program (Hofacker et al. 2002). Hence, from our list of possible causes of a higher synonymous substitution rate in HoxA13a, only an increase of locus specific mutation rate remains.

To further explore this possibility we compared codon usage bias among the HoxA13 paralogs of basal and derived teleosts and outgroup sequences. The rationale for comparing codon bias is that increased mutation rate should move the mutation-selection balance towards lower codon bias, or higher effective number of codons. Figure 2 summarizes the average codon bias for various HoxA13 gene lineages. The numbers represent the effective number of codons, ENC (Wright 1990), that is, higher ENC indicates lower codon bias (Vicario et al. 2007; Salim and Cavalcanti 2008). In the outgroup lineages, the ENC is low and variable, as indicated by the high standard deviation, whereas the HoxA13b paralog clade is relatively uniform with average ENC of about 56 for both the basal lineages (goldeye and eel) and the Ostariophysian lineages (including milkfish and zebrafish). In contrast, the HoxA13a clade shows a variation between the basal lineages (goldeye and shad) and the Ostariophysian lineages, in which the substitution rate is increased. The basal lineages have an ENC of 46.1, whereas the Ostariophysian lineages 60.2 on average. The ENC among the Ostariophysian lineages is also very uniform with a standard deviation of only 1.4. Therefore, the increase in substitution rate of HoxA13a is associated with an increase in ENC, or a decrease in codon bias, consistent with the notion that these genes experience a higher mutation rate. Another possible cause of differences in codon bias could be due to differences in expression level, as genes with higher expression level tend to have higher codon bias or lower ENC (Gouy and Gautier 1982; Sharp et al. 1986; Stenico et al. 1994; Duret and Mouchiroud 1999; Lavner and Kotlar 2005). To examine this possibility we compared the expression levels of the two HoxA-13 paralogs in zebrafish with quantitative PCR.

**EXPRESSION PATTERNS OF PARALOGS DIFFER DURING EMBRYONIC DEVELOPMENT**

Relative quantitative real-time PCR analysis of mRNA present in zebrafish embryos at various developmental stages indicates that HoxA13 paralogs exhibit divergent expression levels during larval development. Because HoxA13 paralogs are expressed in limited number of cells at early stages of development, overall levels of expression are low and difficult to quantify. As a result, HoxA13 expression levels at very early stages of development in zebrafish embryos (6, 12, and 18 hpf) were not detectable. Embryos at 24 hpf, the earliest stage that could be reliably measured, exhibited greater expression of HoxA13b relative to HoxA13a (Fig. 4). However, this pattern was reversed at later developmental stages including 30, 36, and 42 hpf, in which HoxA13a is expressed more than HoxA13b. These data suggest that in early stages of zebrafish larval development, HoxA13b is the predominant paralog expressed. And beginning at approximately 30 hpf and continuing through 54 hpf (48 and 54 hpf not shown), there is a switch in relative expression when HoxA13a becomes the predominant paralog expressed in developing zebrafish embryos. From these data we conclude that the lower estimate of codon bias in HoxA13a, compared to HoxA13b, cannot be explained by a lower expression level of HoxA13a. We conclude that increased mutation rate is the most plausible explanation of both, the lower codon bias and the higher synonymous substitution rate observed in HoxA13a paralogs.

**INSERTIONS AND DELETIONS DEFINED HOPXA13 PARALOG CLADES IN SOME LINEAGES**

In addition to differences in synonymous substitution rates, there were striking differences in insertion and deletion patterns that defined HoxA13a and HoxA13b paralog clades. In HoxA13a, 13 amino acid insertions occurred in the lineage before the origin of the Ostariophysi (Fig. 3, Fig. S2), and after the divergence of the shad, with no deletions compared to ancestral groups. In the evolutionary history of HoxA13b paralogs, five amino acid deletions occurred in the Cypriniformes only, with no insertions relative to
basal groups and HoxA13a sequences. Interestingly, eight of the 13 HoxA13a amino acid insertions and all five HoxA13b amino acid deletions occur in a region spanning only 20 amino acid residues implicating this region as a likely candidate for a novel functional domain in the HOXA13A protein.

**TESTING FOR FUNCTIONAL DIVERGENCE BETWEEN THE HOXA13 PARALOGS**

The fact that we could not detect any evidence for directional selection in either HoxA13 paralog clade can mean one of two things. Either there are no functionally important differences between the two paralog proteins, or the HoxA13a protein has a derived function but the directional selection on the HoxA13a gene could not be detected because of the simultaneous increase in synonymous substitution rate described above, which could deflate the dN/dS ratio. To further test the second possibility, we turned to alternative methods to detect functional differences based on amino acid sequence comparisons. We used the methods described in Gu (Gu 1999, 2001, 2006) as implemented in the program DIVERGE version 1.4.

Neither the method from Gu 1999 nor that of Gu 2001 detected type-I divergence between the two paralogs. We ran the tests with all Ostariophysian species (seven species per paralog clade) as well as only with the four cypriniform species (zebrafish, gold fish, algae eater, and the dojo loach). Similarly neither the Ostariophysian paralog clades, nor the more narrow clade, excluding the milkfish (Chanos chanos), showed any statistical signs of type-II divergence. In contrast, the comparison of the four cypriniform species yielded a strong signal of type-II divergence.

Comparing the cypriniform paralog clades of HoxA13, the coefficient of type-II divergence is $\theta_{II} = 0.215 \pm 0.056$, which is significantly different from zero. Of the 153 alignment sites, 40 have a posterior probability of type-II divergence of larger than 0.5 (Fig. S2). Among those 40 sites, the total expected false positive sites is 7.6, hence within the HoxA13 sequence compared here this method predicts about 32 functionally important amino acid differences between the zebrafish HoxA13a and HoxA13b paralog proteins. The highest posterior probability of type-II divergence is about 0.9 (i.e., either 0.905 or 0.893) and has been detected at 23 alignment sites (Table S4). With the false positive rate of 10%, we would expect only two false positives among these sites. All these alignment sites represent cluster specific differences in all four species of cypriniforms included in this study. Of these differences nine are conservative and 14 are radical amino acid changes according to the classification used by Gu (2006), whereas the differences in the ancestral sequences exhibit 27 conserved and 13 radical amino acid changes. This difference in the proportions between the ancestral and the derived differences is significant ($P = 0.027$, Fisher's exact test). In 18 sites among these 23 highly significant sites, the ancestral amino acid was easily detectable by outgroup comparison (Table S4). Among those 18 sites, the derived amino acid was specific for the HoxA13a clade in 15 cases. This over representation of derived type-II sites in HoxA13a is highly significant ($P = 0.00377$, cumulative binomial probability assuming Pr(HoxA13a derived) = 0.5; $P = 0.00468$, chi-square
test). Therefore, the cypriniform HoxA13a paralog clade exhibits the sequence signature of neofunctionalization.

We conclude that evidence for functional divergence is limited to a relatively derived clade, the HoxA13a sequences of the Cypriniiformes, with a most recent common ancestor of 150 to 180 million years after the teleost specific whole genome duplication (Inoue et al. 2005). Furthermore, HoxA13a paralog has the signature of neofunctionalization, that is, has a preponderance of derived highly significant type-II amino acids. Taken together, these findings imply that neofunctionalization occurred at least 150 Mio years after the gene duplication, and therefore is unlikely to be the cause for the initial retention of the duplicate HoxA13 genes.

FUNCTIONAL PERFORMANCE OF HOXA13 PARALOGS IN ZEBRAFISH

Transcription factor function of zebrafish HoxA13 paralogs is conserved

The transcriptional activity of Hox proteins can be modulated by interaction with other proteins, many of which are expressed in a cell type specific manner (Zappavigna and Mavilio 1994; Saleh et al. 2000; Yang et al. 2000; Kobayashi et al. 2003; Lu et al. 2003; Bondos et al. 2004; Shen et al. 2004; Luke et al. 2006). To compare the unmodulated transcriptional activity of Hox proteins, which are expressed in different cell types, it is necessary to choose a cell line that is equally foreign to both proteins. We therefore employed the GAL4/UAS system in COS-1 cells, which are derived from African green monkey kidneys, to determine the transcriptional activity of the cypriniform HoxA13 paralogs. In this system, the HoxA13 paralogs are expressed as fusion proteins with the GAL4 DNA binding domain. The influence of these fusion proteins on the expression of a reporter gene, in our case luciferase, was tested according to Schnabel and Abate-Shen (1996). In this system, HoxA13a and HoxA13b function as mild transcriptional repressors (Fig. 5). We conclude that the asymmetric divergence in zebrafish HoxA13a and HoxA13b paralogs has not affected the unmodulated transcription factor function. Together with the fact that the homeodomain of the paralogous proteins are identical (with the exception of one Q/H difference at the variable amino acid site 37 of the homeodomain) our results suggest that any functional differences are likely to be due to protein–protein interactions with other transcription factors or protein–RNA interactions rather than changes in the unmodulated transcriptional function or DNA-binding specificity.

Expression of HoxA13 zebrafish paralogs via in situ hybridization at early stages of development

Expression patterns were difficult to visualize via whole mount in situ hybridization at stages before 24 hpf. We consistently found expression of both paralogs in the developing tails and pectoral fins from 24 to 48 hpf similar to that reported for HoxA13b (Géraudie and Borday Birraux 2003; Thisse and Thisse 2005). In addition, we observed slight fields of expression in some individuals, but the results were too inconsistent for assessing bona fide differences between paralogs. Therefore, we focused on a more direct method of evaluating functional divergence using a knockdown and rescue approach.

Functional divergence of HoxA13 paralogs assessed by knockdown and rescue

To test for functional differences between HOXA13 paralogs, we used paralog specific Morpholinos (MO) to block mRNA translation. Paralog specific antisense MOs were designed to target the paralog specific 5' UTR region of the mRNA. Injections of these MOs into 1–2 cell stage embryos in over 500 individuals from multiple experiments resulted a variety of consistently observed phenotypic effects (Table 2). Phenotypic effects that are similar in the HoxA13a and HoxA13b knockdowns clearly suggest that there is some degree of redundancy afforded by maintaining both paralogs, although it is conceivable that some of the aberrant phenotypes may be partially due to secondary effects or to general MO effects (Robu et al. 2007). Both knockdowns resulted in a phenotype with a characteristic bent tail, but the HoxA13a knockdown resulted in a more anterior location of the effect. This effect is consistent with the expression of both paralogs in the growing tail tip (see above). However, there are several phenotypic effects that are strikingly different between the knockdowns, providing evidence for functional divergence in zebrafish HoxA13 paralogs (Table 2). A major difference is the absence (or severe reduction) of the YSE in the HoxA13a knockdown (Fig. 6), seen at 32 hpf in >80% of all injected embryos. The MO designed to knockdown
**Table 2.** Summary of effects due to knockdown of *HoxA13a* and *HoxA13b* via paralog specific morpholinos (MO) injected into zebrafish embryos at the in 1–2 cell stage. Note that the yolk sac extension (YSE) is absent or greatly reduced in the *HoxA13a* knockdown morphant.

<table>
<thead>
<tr>
<th>Feature</th>
<th><em>HoxA13a</em> MO</th>
<th><em>HoxA13b</em> MO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yolk sac extension</td>
<td>Greatly reduced or absent</td>
<td>Present</td>
</tr>
<tr>
<td>Bent tails</td>
<td>Present, more anterior</td>
<td>Present, more posterior</td>
</tr>
<tr>
<td>Caudal fin fold</td>
<td>Absent</td>
<td>Present</td>
</tr>
<tr>
<td>Pectoral fin at 48 hpf</td>
<td>Delayed; only fin primordia</td>
<td>Delayed; only fin primordia</td>
</tr>
<tr>
<td>Pectoral fin at 106 hpf</td>
<td>Delayed; small fin buds</td>
<td>Delayed; small fin buds</td>
</tr>
<tr>
<td>Cardiac edema at 72 hpf</td>
<td>Marked</td>
<td>Marked</td>
</tr>
<tr>
<td>Notochord</td>
<td>Disorganized caudally</td>
<td>Disorganized caudally</td>
</tr>
<tr>
<td>Body axis extension</td>
<td>Normal</td>
<td>Shorter</td>
</tr>
<tr>
<td>Body width</td>
<td>Reduced</td>
<td>Normal</td>
</tr>
<tr>
<td>Microcephaly</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Pigmentation pattern</td>
<td>Melanophores lighter in color and less organized</td>
<td>Melanophores less organized</td>
</tr>
<tr>
<td>Circulation at 106 hpf</td>
<td>Greatly reduced</td>
<td>Somewhat reduced</td>
</tr>
<tr>
<td>Neuromasts at 106 hpf</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>

*HoxA13b* did not have an effect on the development of the YSE. Other differences include development of the caudal fin fold, body width, pigmentation, and circulation rates. Most importantly, the effect of the *HoxA13a* MO on formation of the YSE is significant and suggests that this paralog may function in the formation of endodermal YSL, or formation of the blood stem cell compartment (Hammerschmidt et al. 1996; Braat et al. 1999; Leonard A. D’Amico 2001; This and Zon 2002). This is notable because the YSE is a novel developmental character that is associated with zebrafish and some related taxa (Virta and Cooper 2005, 2009) within the Ostariophysi.

To test for the specificity of the *HoxA13a* MO on YSE, we performed rescue experiments by co-injecting capped synthetic mRNA together with the *HoxA13a* MO and scored for retention of the YSE (Table 3, Fig. 7). The rescue experiment with the *HoxA13a* mRNA resulted in partial or full retention of the YSE in the vast majority of embryos when compared to the MO-only controls (98% vs. 32%, respectively). Additionally, the number of embryos that completely lacked the YSE was much lower in the RNA-co-injected embryos relative to the MO-only controls (2.4% vs. 68.4%, respectively). We thus conclude that partial rescue of the YSE was achieved via the addition of *HoxA13a* mRNA and that the loss of the YSE is a specific effect of *HoxA13a* knockdown. In the rescue experiment with the *HoxA13b* mRNA, most cases showed no rescue (51%). However, we did observe partial rescue in 19% of the cases, and severe abnormalities in 28%. It may be that when overexpressed in tissues that normally express *HoxA13a* but not *HoxA13b*, *HoxA13b* can provide partial rescue, but further experiments are needed to confirm this interpretation. Regardless, rescue of *HoxA13a* MO by *HoxA13b* was far less reliable than by *HoxA13a*.

**Discussion**

In this article, we investigated the sequence and functional divergence of duplicated *HoxA13* paralogs in zebrafish and related...
Table 3. Rescue of HoxA13a-morpholino treated embryos with HoxA13a capped mRNA. Wildtype zebrafish embryos were injected at the one- or two-cell stage with HoxA13a MO alone or with Hoxa13a MO and synthetic mRNA of HoxA13a. Embryos were scored for the appearance of their yolk sac extension at 32 hpf. Data are taken from two independent experiments. Figures in parentheses represent percentages.

<table>
<thead>
<tr>
<th>YSE phenotype at 32 hpf</th>
<th>Sham-injected</th>
<th>MO only</th>
<th>MO plus HoxA13a mRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full YSE</td>
<td>72 (98.6)</td>
<td>6 (10.5)</td>
<td>20 (24.4)</td>
</tr>
<tr>
<td>Reduced YSE</td>
<td>1 (1.4)</td>
<td>12 (21.1)</td>
<td>60 (73.2)</td>
</tr>
<tr>
<td>No YSE</td>
<td>0</td>
<td>39 (68.4)</td>
<td>2 (2.4)</td>
</tr>
<tr>
<td>Total</td>
<td>73</td>
<td>57</td>
<td>82</td>
</tr>
</tbody>
</table>

taxa. Specifically we sought to understand the sequence of events that led from gene duplication to two fixed and functionally differentiated paralog genes in the zebrafish genome. Comparison of paralogs between distantly related species cannot resolve how or when in phylogeny the differences arise, nor what evolutionary forces have contributed to (or are responsible for) such differences (Lynch and Katju 2004). For instance, the patterns of sequence divergence of HoxA13 paralogs in zebrafish and fugu are different, even though they arose from the same genome duplication. In the zebrafish lineage, the HoxA13a paralog is more diverged whereas in fugu the HoxA13b paralog is more derived (Wagner et al. 2005). Clearly, the majority of differences between zebrafish and fugu arose after divergence from their most recent common ancestor. Only a more extensive taxon sampling, as we present here, can resolve the question of when and how the paralogs arose and diverged.

**Molecular Evolution of HOXA13 Paralogs in Zebrafish and Related Taxa**

Both HoxA13 paralogs have been maintained in zebrafish, and in this study, we present evidence for retention and maintenance of HoxA13 paralogs in the basal teleost goldeye, and representatives from four of the five major lineages in the Ostariophysi. These paralogs are products of a whole genome duplication that occurred prior to the teleost radiation (Hoegg et al. 2004; de Souza et al. 2005; Crow et al. 2006; Salaneck et al. 2008), and exhibit significant differences in substitution rates and patterns in respective lineages. This suggests that there may be different evolutionary forces acting on HoxA13 paralogs in various lineages leading to functional divergence.

The whole genome duplication that occurred in ray-finned fishes is well supported by evidence from genomic studies in zebrafish (Amores et al. 1998; Taylor et al. 2003), medaka (Naruse et al. 2000), and pufferfish (Christoffels et al. 2004; Jaillon et al. 2004). The timing of the duplication occurred shortly before the origin of teleosts, as indicated by gene genealogies of Hox genes and many other loci from a variety of focused studies in teleosts (Hawkins et al. 2000; Hoegg et al. 2004; de Souza et al. 2005; Crow et al. 2006; Schweitzer et al. 2006; Cardoso et al. 2007; Hurley et al. 2007; Salaneck et al. 2008). The established history of the teleost specific genome duplication would predict a specific topology for any genealogy with two resolved paralog clades after the duplication. The estimated topology of HoxA13 does not conform to this expectation, yet is informative because it clearly demonstrates that rates of evolution between basal teleosts are different from those of more derived groups. More importantly, rates of evolution are asymmetric between Ostariophysan HoxA13 paralogs in derived clades, with faster rates of evolution and the tendency to accumulate insertions in the HoxA13a clade, and slower rates of evolution and amino acid deletions in the HoxA13b clade.

**Figure 7. Rescue of HoxA13a morphant with HoxA13a capped mRNA.** Wildtype zebrafish embryos were injected at the 1–2 cell stage with HoxA13a morpholino alone (top left), sham buffer (bottom left), or with Hoxa13a MO and synthetic mRNA to HoxA13a (right panel). Note various degrees of YSE restoration in the rescued embryos.
Increased rates of evolution and divergence after duplication are well-documented phenomena (Lynch and Conery 2000; Kondrashov et al. 2002; Conant and Wagner 2003; Wagner et al. 2005). However, one criticism of these observations is that little is known about how and when the asymmetric divergence takes place because taxon sampling is limited (Lynch and Katju 2004). Our data show that the insertions in HoxA13a paralogs were acquired after the divergence of the Clupeomorphs (shad) but before the origin of the Ostariophysi. The deleterions in the HoxA13b paralogs occurred after the divergence the Siluriform/Gnathostomata clade from the Cypriniformes (including zebrafish). Another important result of our analyses is that the synonymous substitution rate for HoxA13a increased dramatically after the divergence of Clupeomorpha from the Ostariophysi. Branch lengths in Ostariophysans HoxA13a paralogs are longer in all but one comparison with HoxA13b, and 71% are significant. The onset of the asymmetric divergence of the HoxA13 paralogs occurred about 40 to 80 Mio years after the gene duplication event rather than immediately following it, based on the divergence time estimates published by Inoue et al. (2005). Finally, we found that sequence evidence for type-II functional divergence is observed only in cypriniform taxa in our study, which diverged more than 100 Mio years after the duplication event.

Overall, we find no evidence for episodes of positive selection in the evolutionary history of the HoxA13 gene in Ostariophysans. In fact, it was surprising to learn that the predominant process driving the asymmetry in HoxA13 paralog evolution was an increase in synonymous substitution rates in HoxA13a relative to HoxA13b, even though there was an increase in nonsynonymous substitutions as well. Although there could be specific amino acids under positive selection, the signature of selection was probably swamped by the overwhelming rate increase of synonymous substitutions, and thus the power to detect positive selection in specific lineages is limited.

We consider the possibility that selection may be acting on characters other than amino acid substitution to increase the synonymous substitution rate in the HoxA13a paralogs of ostariophysans. It is possible for selection to favor synonymous substitutions to modify the GC content of a locus, to satisfy optimal codon usage (i.e., codon bias), or to favor a locus-specific increased mutation rate (e.g., site specific inhibition of repair enzymes during DNA replication). We found no evidence for directional selection on synonymous substitution rates in HoxA13a sequences relative to HoxA13b sequences in terms of codon usage bias or GC content in basal or derived clades. Another scenario is that there is an increased mutation rate in both paralog clades, but there are constraints placed on the HoxA13b paralogs such as conserved secondary structure in the mRNA or antisense transcripts. However, we found no evidence for mRNA secondary structure in zebrafish HoxA13b paralog sequences. Therefore, we are left to conclude that the mutation rate is probably higher in the Ostariophysan lineage of HoxA13a paralogs. This interpretation is corroborated by the observation that the faster evolving HoxA13a genes have a lower codon bias and are more strongly expressed than HoxA13b in zebrafish development.

The idea that mutation is not only randomly introduced into the genome, but that some sites may exhibit higher mutation rates than others as sculpted by natural selection was introduced by Lynn Caporale (Caporale 2003a,b). Caporale illustrates this focused hypermutation with examples from the protein coat encoding genes of bacteria and molecules involved in the vertebrate immune response. Site-specific hyperdivergence also has been observed between exons of genes encoding conotoxins in marine snails of the genus Conus (Woodward et al. 1990; Olivera et al. 1999). In the case of conotoxins, three regions in conotoxin genes, separated by introns, exhibit strikingly different substitution rates. The conserved regions are thought to play a role in protein folding, whereas the mature venom peptide is encoded by the hypervariable region.

The increased mutation rate (and associated asymmetry) in zebrafish appears to be specific to the HoxA13a locus. The idea that mutation rates can be lineage- and locus-specific for a duplicated Hox gene is further supported by several observations. First, rate asymmetry was not observed in nonsynonymous substitution rates of zebrafish HoxA11 paralogs (Wagner et al. 2005). Second, there was no significant difference in substitution rates in basal teleosts compared to basal ray-finned fish (illustrated in Fig. 1, P-values for branch length comparisons not shown). Finally, an alternative pattern has been observed in the pufferfish, Takifugu rubripes, in which it is the HoxAb paralog that exhibits asymmetry and accelerated rates of evolution. In this case, rate change is associated with a whole Hox cluster because all five fugu HoxAb genes exhibit greater rates of evolution, three of which were significantly faster (Wagner et al. 2005).

**FUNCTIONAL DIVERGENCE IN ZEBRAFISH HOXA13 PARALOGS**

Asymmetric divergence of duplicate genes has been taken as suggestive evidence for the acquisition of a novel function by the paralog with the higher rate of divergence (Conant and Wagner 2003; Kellis et al. 2004). The HoxA13b sequences are more similar to the HoxA13 sequences of basal taxa that diverged prior to the whole genome duplication than to HoxA13a sequences. Clearly, the HoxA13a paralog has a more dynamic post duplication history than HoxA13b, with evidence for hypermutability resulting in increased rates of $d_N$ and $d_S$, as well as amino acid insertions which could implicate this gene in derived biological roles. To test this, we performed morpholino knockdown experiments with paralog specific MOs. Indeed, we found evidence for functional divergence between paralogs, suggesting
that changes in the molecular evolutionary history of the \textit{HoxA13a} paralog is associated with the development of a novel character (YSE).

In sequence-based tests for functional divergence, we found that cypriniforms exhibit a strong signal for type-II functional divergence (i.e., fixed differences between corresponding amino acid sites among paralogs) but no evidence for divergence of more inclusive clades such as the Otophyi or the Ostariophysi. A statistical estimate suggests that there are at least 32 functionally important amino acid differences between the zebrafish \textit{HoxA13} paralogs. Furthermore, among the highly significant type-II sites (posterior probability of type-II divergence of about 90%) the majority of derived amino acid residues are found in \textit{HoxA13a}. We take this as evidence that \textit{HoxA13a} has acquired a novel function in the stem lineage of Cypriniformes. This conclusion is confirmed with a knockdown experiment with paralog specific MO. Only the \textit{HoxA13a}MO prevented the development of the YSE in zebrafish embryos, suggesting that YSE development requires the \textit{HoxA13a} paralog.

**FUNCTIONAL DIFFERENCES BETWEEN THE \textit{HOXA13A} AND \textit{HOXA13B} PARALOGS WITH RESPECT TO THE YSE—A DERIVED PHENOTYPE**

Formation of the YSE in zebrafish has been described as one of the most dramatic morphogenetic events during development of the YSL, which begins around the 15-somite stage (Cooper and Virta 2007). A sustained contractile event reshapes the posterior region of the YSL, forcing yolk to flow both anteriorly and posteriorly as the contractile band constricts in a ventral-to-dorsal direction. The YSE appears to be a conserved character in Cypriniformes (Virta and Cooper 2005), and embryos from another Ostariophysan group, the Characiformes, have recently been described as exhibiting a YSE (Ninhaus-Silveira et al. 2006, 2007). We found no evidence in the literature of YSE formation in the Clupeomorpha (sister group of the Ostariophysi). The yolk extension in the milkfish, a member of the most basal Ostariophysan lineage, is described as being extended to form an elongated morphology, or cylindrical teardrop (Virta, pers. comm., and illustrated in Chaudhuri et al., 1978). Results from the knockdown experiments using paralog specific \textit{HoxA13} MOs show that \textit{HoxA13a}, the more diverged paralog, may be necessary for development of the YSE in some Ostariophysan fishes. This phenotype can be rescued with injection of \textit{HoxA13a} mRNA. These results provide clear evidence for functional nonequivalence between \textit{HoxA13} paralogs in zebrafish. We propose that this may be an example of an evolutionary novelty (the YSE) caused by neofunctionalization of a duplicated gene. Based on these data, we propose that the development of the YSE may be an Otophyi invention (Ostariophysan lineages except the Gonorynchiformes), but taxon sampling of this character is currently too sparse to allow firm inferences about the possibility of parallel derivations of a YSE in various ostariophysan lineages. Type II functional divergence seems to be limited to cypriniforms, which would speak for a Cypriniform derivation of the zebrafish YSE, and parallel derivation of similar phenotypes in other ostariophysan groups. Another possibility is that the protein changes found by Type II divergence tests relate to the canalization of the phenotype rather than to its origin.

The exact functional significance of the YSE is not clear, but results from a similar phenotype suggest a role in early mesodermal patterning; the \textit{HoxA13a} morphant phenotype is similar to \textit{kugelig} (\textit{kgg}). \textit{Kgg} abolishes both YSE and blood-island development and is known as an upstream regulator of certain Hox genes (Davidson et al. 2003). It will thus be interesting to investigate whether similar processes are affected in the \textit{HoxA13a} morphant as in the \textit{kgg} mutant.

**SUMMARY**

One of the most intriguing and timely questions in evolutionary biology today is to understand the types of evolutionary changes responsible for generating phenotypic diversity. Here, we present evidence that hypermutability in one paralog of a duplicated Hox gene, and later type-II functional amino acid divergence is associated with the evolution and development of a novel feature in a group of fishes related to zebrafish. We are able to infer that changes at the molecular level are apomorphic by expanded taxon sampling. In addition, we clearly demonstrate that although there are some redundant phenotypic effects associated with both paralogs, development of the YSE feature is specified by one paralog—the more divergent \textit{HoxA13a}. This paralog is evolving comparatively rapidly in specific lineages resulting in dramatic structural differences in the coding sequences of duplicated \textit{HoxA13} genes in zebrafish.

Based on these data, we propose the following scenario for the evolutionary history of the duplicated \textit{HoxA13} genes in zebrafish and relatives (Fig. 8).

(1) There was a latent phase in evolutionary history of the \textit{HoxA13} paralogs prior to divergence of the Ostariophysan lineages.

(2) After the divergence of shad from the Ostariophysi, rates of both \(d_s\) and \(d_\text{N}\) increased in \textit{HoxA13a}, likely due to a paralog specific increase in mutation rate, and 13 amino acid insertions were acquired.

(3) In the stem lineage of the Cypriniformes, the \textit{HoxA13a} paralog acquired functionally important amino acid changes that remain fixed in the cypriniform species examined here.

(4) Molecular changes in \textit{HoxA13a} are associated with development of a novel feature, the YSE.
EVOLUTION

Figure 8. Proposed evolutionary scenario of the molecular evolution of the duplicated HoxA13 genes in zebrafish and relatives. We suggest that there was a latent phase in HoxA13 paralog divergence for tens of millions of years, followed by a phase of increased mutation rate in HoxA13a, leading to the origin of a novel functionality in HoxA13a in the stem lineage of the cypriniforms.

ACKNOWLEDGMENTS

Special thanks to Kent Dunlap for providing a tissue sample of the electric knifefish Apteronotus leptorhynchus, and to Guillermo Orti for providing a tissue sample of the milkfish Chanos chanos. Sarah Oneto and Emily Edenburn helped produce three HoxA13b sequences as participants in the “Research Internships for Secondary Science Teachers” program at Yale University in summer 2006. The pM2, pGL25xGal, and pCMV-βgal plasmids were gifts from the Abate-Shen group. This work was supported by National Science Foundation grants IOS-0321470 (to GPW) and IOS-0321461 & MCB-0719558 (to CTA). JJR was awarded a fellowship from the “Research Internships for Secondary Science Teachers” program at Edenburn.

LITERATURE CITED


Supporting Information

The following supporting information is available for this article:

**Figure S1.** Saturation curves of *HoxA13* paralog sequences in Ostariophyans indicating that transitions in *HoxA13* sequences are not saturated, (i.e., increasing linearly at approximately the same rate as transversions).

**Figure S2.** Posterior probability of Type II divergence in cypriniform *HoxA13* paralogs.

**Table S1.** Sequences for primers, probes and morpholinos (MO) used for amplification of *HoxA13* paralogs in ray-finned fishes, and detection or knockdown in zebrafish.

**Table S2.** Model selection for aligned data sets of three *HoxA13* lineages including basal lineages (bichir, gar, bowfin), and Ostariophysan *HoxA13a* and *HoxA13b* paralog clades.

**Table S3.** Pairwise relative rate comparison between *HoxA13* paralogs in Ostariophyans using the bowfin (Aca) as the outgroup.

**Table S4.** Amino acid alignment positions with type II functional divergence between cypriniform *HoxA13a* and *HoxA13b*.

Supporting Information may be found in the online version of this article.

(This link will take you to the article abstract).

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Additional results and discussion can be found in a document at http://www.repository.naturalis.nl/record/289893.