J Mol Evol (2004) 59:190–203
DOI: 10.1007/s00239-004-2613-z

JOURNAL OF MOLECULAR EVOLUTION

Phylogenetic Timing of the Fish-Specific Genome Duplication Correlates with the Diversification of Teleost Fish

Simone Hoegg, Henner Brinkmann,* John S. Taylor,† Axel Meyer

Department of Biology, University of Konstanz, 78457 Konstanz, Germany

Received: 6 August 2003 / Accepted: 13 February 2004 [*Reviewing Editor*: Martin Kreitman]

Abstract. For many genes, ray-finned fish (Actinopterygii) have two paralogous copies, where only one ortholog is present in tetrapods. The discovery of an additional, almost-complete set of Hox clusters in teleosts (zebrafish, pufferfish, medaka, and cichlid) but not in basal actinopterygian lineages (*Polypterus*) led to the formulation of the fish-specific genome duplication hypothesis. The phylogenetic timing of this genome duplication during the evolution of ray-finned fish is unknown, since only a few species of basal fish lineages have been investigated so far. In this study, three nuclear genes (*fzd8*, *sox11*, *tyrosinase*) were sequenced from sturgeons (Acipenseriformes), gars (Semionotiformes), bony tongues (Osteoglossomorpha), and a tenpounder (Elopomorpha). For these three genes, two copies have been described previously teleosts (e.g., zebrafish, pufferfish), but only one orthologous copy is found in tetrapods. Individual gene trees for these three genes and a concatenated dataset support the hypothesis that the fish-specific genome duplication event took place after the split of the Acipenseriformes and the Semionotiformes from the lineage leading to teleost fish but before the divergence of Osteoglossiformes. If these three genes were duplicated during the proposed fish-specific genome duplication event, then this event separates the species-poor early-branching lineages from the species-rich teleost lineage. The additional number of genes resulting from this event might have facilitated the evolutionary radiation and the phenotypic diversification of the teleost fish.

Key words: Genome duplication — Gene duplication — Actinopterygii — *sox* — *fzd* — *tyrosinase*

Introduction

Most increases in gene numbers occur through many independent tandem duplication events, yet rare entire genome duplications appear to have played a major role during the evolution of genomic and possibly phenotypic complexity (e.g., Ohno 1970; reviewed in Meyer and Van de Peer 2003). Studies first on genome size (Ohno 1970) and later on various gene families supported the idea that duplications of whole genomes had an important impact, in particular, on the evolution of vertebrates (Ohno 1970; Spring 1997). The most prominent example of a correlation between genomic and phenotypic complexity is the clusters of Hox genes. All tetrapods have four Hox clusters with a total of 39 genes (Acampora et al. 1989; Graham et al. 1989; Harvey et al. 1986), but the evolutionarily more basal and morphologically rather simple cephalochordate amphioxus (*Branchiostoma floridae*) has a single cluster with only 14 Hox genes (Garcia-Fernandez and Holland 1994). These data are consistent with the hypothesis of two rounds of genome duplications

Present addresses:
*Département de biochimie, Université de Montreal, Montreal, QC H3C3J7, Canada
†Department of Biology, University of Victoria, Victoria, BC V8W3N5, Canada
Correspondence to: Axel Meyer; *email:* axel.meyer@uni-konstanz.de

within the vertebrate lineage (Spring 1997; reviewed in Meyer and Van de Peer 2003).

Genome sequencing projects provide data about the evolution of gene numbers and the diversification of gene families when analyzed in a phylogenetic context. Based on data from human (*Homo sapiens*), mouse (*Mus musculus*), chicken (*Gallus gallus*), nematode (*Caenorhabditis elegans*), fly (*Drosophila melanogaster*), thale cress (*Arabidopsis thaliana*), rice (*Oryza sativa*), and yeast (*Saccharomyces cerevisiae*), the rate of gene duplications was estimated at about 0.01 duplication per gene per million years, which is of the same order of magnitude as the mutation rate per nucleotide site (Lynch and Conery 2000, 2003). Based on this estimate, one might expect that the number of duplicated genes in genomes would be much higher. But in many cases, one of the duplicated copies acquires mutations quickly, leading to its inactivation and loss within a short time (Li 1980). Lynch and Conery (2000) and Lynch (2002) estimate that the half-life of a duplicated gene is only of the order of 4.0 million years (Lynch and Conery 2003), and therefore, the increase in the number of genes in genomes due to small-scale tandem duplications is counteracted by a rather short half-life and a relatively high rate of gene loss. Mutations can also result in functional changes and then the two copies might no longer be identical and redundant in function. Neofunctionalization (Ohno 1970; Sidow 1996) and subfunctionalization (Force et al. 1999) are processes that would be expected to promote the retention of duplicated genes.

For many gene families, two paralogous copies are found in zebrafish and pufferfish, where only one ortholog is present in tetrapods (Wittbrodt et al. 1998). The discovery of larger gene families in fish has led to the formulation of the fish-specific genome duplication hypothesis (Amores et al. 1998; Wittbrodt et al. 1998; Ohno 1999; Taylor et al. 2001a, b, 2003; reviewed in Meyer and Van de Peer 2003; but see Robinson-Rechavi et al. 2001), which states that during the evolution of vertebrates, a duplication of the entire genome occurred in the fish lineage but not in the lineage leading to land vertebrates (tetrapods). In zebrafish (*Danio rerio*) and medaka (*Oryzias latipes*), seven Hox clusters were identified: two HoxA, HoxB, and HoxD clusters and one HoxC cluster (Amores et al. 1998; Naruse et al. 2000). In the cichlid fish genome, six Hox clusters were found (Màlaga-Trillo and Meyer 2001) and at least seven clusters in the genomes of pufferfish (Aparicio et al. 2002; Amores et al. 2004). These and other findings provided support for the idea of an additional genome duplication during the evolution of the ray-finned fish (actinopterygians) (Amores et al. 1998; Wittbrodt et al. 1998). More evidence for this fish-specific duplication event comes from other genes that have two copies in modern fish, such as *Danio rerio* or *Takifugu rubripes*, but only one copy in tetrapods (Chiang et al. 2001; Kao and Lee 2002; Lister et al. 2001; Merritt and Quattro 2001). For a large number of duplicated genes in fish, the two copies are located on different chromosomes. Sometimes, even hundreds of millions of years after the duplication, synteny between two chromosomes can be found (Gates et al. 1999; Barbazuk et al. 2000; Postlethwait et al. 2000).

Phylogenetic analyses of 27 gene families showed that in 15 cases the two paralogous zebrafish copies are more closely related to each other than to the tetrapod orthologs (Taylor et al. 200la). The incorporation of data from the *Takifugu rubripes* genome sequencing project confirms, for 42 of a total of 49 genes, that the duplication is not specific to the lineage leading to *Danio rerio* (Taylor et al. 2003). The recent discovery of seven Hox clusters in the pufferfish genome (Aparicio et al. 2002; Amores et al. 2004) also indicates that the duplication event took place before the last common ancestor of the Neoteleostei (*Takifugu rubripes, Spheroides nephelus*) and Ostariophysii (*Danio rerio*).

However, an increased number of duplicated genes in the ray-finned fish lineage is in itself not sufficient evidence for a fish-specific genome duplication, unless it can be shown that all these ''extra'' fish genes originated at the same time and are not the result of many independent tandem gene duplications or several chromosomal duplications during the evolution of fish. Mapping data for many of those duplicated genes show that they in fact do not result from tandem duplications (Taylor et al. 2001, 2003). Information about gene copy numbers from basal fish lineages is required for genes that are found in two copies in the derived fish lineages, *Danio rerio* (zebrafish) and *Takifugu rubripes* (pufferfish), which have been investigated so far in the most detail, in order to determine the timing of the evolutionary origin of the gene paralogs.

The class Actinopterygii includes about 23,700 species (Nelson 1994); the vast majority of them (23,637 species [Nelson 1994]) belongs to the division Teleostei (teleost fish). The older groups, namely, Polypteriformes (bichirs), Acipenseriformes (containing the families Acipenseridae [sturgeons] and Polyodontidae [paddlefish]), Semionotiformes (gars), and Amiiformes (bowfin), consist of only a few extant species. Most members of these basal actinopterygian lineages are considered to be living fossils (Eldredge and Stanley 1984), because their morphology remained unchanged over very long periods of time. Pufferfish and zebrafish shared a last common ancestor approximately 284–296 million years ago (mya) based on a calibration from molecular data (Kumazawa et al. 1999), and the split between Sarcopterygii and Actinopterygii is about 450 million

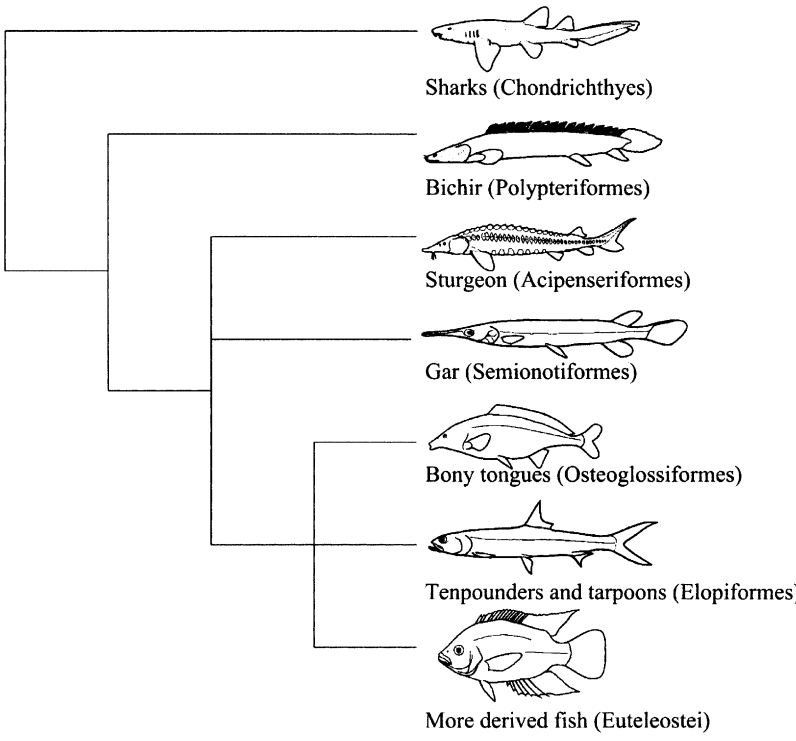

Fig. 1. Consensus of phylogenetic relationships among basal actinopterygian fish. Simplified tree from Inoue et al. (2003).

years old (Kumar and Hedges 1998; Hedges and Kumar 2003). The average age for paralogous genes was estimated to be between 300 and 450 mya (Taylor et al. 200la), a time window that would include the origin of all basal orders of ray-finned fish and, therefore, gives no indication of the relative position of the duplication event.

Although several studies on duplicated genes have been performed, it is still not known when during the evolution of fish the fish-specific genome duplication happened and which, if any, living fish lineages originated before this large-scale genomic event. For a correct positioning of the duplication event, a robust phylogeny of these early diverging fish lineages is needed. The phylogenetic relationships among the basal Actinopterygii are still controversial, especially among palaeontologists, and the first molecular approaches did not provide clear answers as well (Le et al. 1993; Normark et al. 1991). Palaeontological and comparative morphological data provide evidence for *Polypterus* being the most basal of the actinopterygians, with the Acipenseriformes branching off next (Grande and Bemis 1996; Noack et al. 1996). The relative position of the bowfin and the gar with respect to the Teleostei based on palaeontological/morphological ground is still debated (Arratia 2001). Early molecular data point toward the monophyly of Holostei with gar and bowfin as a sister group of the Teleostei (Le et al. 1993; Normark et al. 1991). Some paleontological analyses rather support paraphyly of the Holostei, with *Amia calva* being more closely related to the Teleostei (Patterson 1973; Schultze and Wiley 1984; Wiley and Schultze 1984). Only recent molecular approaches using whole mitochondrial genome sequences produced highly supported trees, with the Osteoglossomorpha (bony tongues) as the most basal teleost groups and the Elopomorpha (tenpounders and eels) representing a more recent lineage (Inoue et al. 2001). The Polypteriformes are positioned at the base of the actinopterygians, while the Acipenseriformes form a monophyletic group with the Semionotiformes and Amiiformes, even though likelihood ratio tests could not reject alternative topologies (Inoue et al. 2003) (Fig. 1).

In this study, we used PCR to look in sturgeon, gar, elephantnose fish (a representative of the osteoglossomorphs), and tenpounder for orthologs of genes known to have been duplicated in fish before the divergence of zebrafish and pufferfish. The cichlids (*Oreochromis niloticus* and *Amphilophus citrinellum*) were selected as additional Neoteleostei species, a group that is phylogenetically younger than the Ostariophysii, represented by the zebrafish, and therefore also belongs to the clade that is already known to have experienced the genome duplication. We determined DNA sequences of three protein–encoding nuclear genes *fzd8, sox11* and, *tyrosinase*.

Fzd8 belongs to the family of Wnt receptors with seven transmembrane domains and an extracellular cysteine-rich domain (CRD) at the amino terminus (Wang et al. 1996). The gene is made up of a single open reading frame, which codes for approximately 580 amino acids. Phylogenetic analyses together with

Table 1. List of PCR primers used in this study

Primer name	Sequence 5′ → 3′	Length	aa motif
FZD8.uni.F190	GGY TAY AAY TAC ACC TAC ATC CC	23	GYNYTYMP
FZD8.uni.F1265	ATG GCS AGC KCC ATC TGG TGG	21	MASSIWW
FZD8.uni.R1265	A CCA GAT GGM GCT WGC CAT BCC	22	GMASSIW
FZD8a.uni.R2090	GG ACA AWG GCA TCT GCT TGG	20	KQMPLS
FZD8b.uni.R1915	CGA YCK CCA SGT CAG TCC C	19	GLTWRS
Sox11.uni.F270	CK CCR GAC ATG CAC AAC GC	19	SPDMHNA
Sox11.uni.F180	GAC TGG T TGC AAR ACA GCM AC	20	DWCKTAT
Sox11.uni.R1200	G CAA GTC GTC SGI GTC YTC GC	21	EDSDDLL
Sox11.uni.R1390	TC IGG IGT GCA ATA GTC YGG	20	PDYCTPE
Sox11a.uni.F5′E	TG CAG CAM ACS GAC AAC AGC	21	VQQTDNS
Sox11b.unLF5′E	G GTG CAG CAR ACR GAR CAI AG	21	VQHTEQ/H
Sox11.uni.3′E	AA IAC CAR RTC IGA AAA GTT MGC	23	ANFSDLVF
Soxbox.Sim.dir	ATG AAY GCI TTY ATG GTI TGG	21	MNAFMVW
Soxbox.Sim.rev	GG YCK RTA YTT GTA GTC IGG	20	PDYKYRP
Tyr.F40 (aa)	AAR GAR TGY RGY CCI GTI TGG	23	KECCPVW
Tyr.F25 (aa)	GI CAG TTY CCY MGI GCY TG	19	QFPRA
Tyr.RSPA	GA IGA GAA RAA RGA IGC TGG GCT	23	SPASFFSS
Tyr.RCTD	CC ICC CAW IAR YTC ATC WGT GCA	23	CTDELM/FGG
Tyn.RWDW	TS IGC ATC YCK CCA RTC CCA	20	WDWRDA

Note. uni, universal; F, forward; R, reverse. The numbers refer to the position in the nucleotide (*fzd8*/*sox11*) or amino acid (*tyrosinase*) alignment. The *tyrosinase* reverse primers are labeled by their binding sequences.

frizzled genes from human, mouse, and frog showed that the two zebrafish paralogs are more closely related to each other than to any other sequence of the dataset (Van de Peer et al. 2002b). Database searches of the available nearly complete genome sequences also revealed two copies of this gene in *Takifugu rubripes* (Taylor et al. 2003).

Sox11 belongs to the SOX family of transcription factors, which are characterized by a 79-amino acid motif, the HMG (high mobility group) box that was first discovered in the mammalian testis-determining factor SRY (Bowles et al. 2000). In zebrafish, two *sox11* paralogues, *sox11a* and *sox11b*, were found and mapped to linkage groups 17 and 20, respectively (Rimini et al. 1999); the possibility of a recent tandem duplication is therefore excluded. Paralogs of other genes have already been mapped to these two linkage groups as *bmp2a*/*bmp2b* and *snap25b*/*snap25a* (Martinez-Barbera et al. 1997; Risinger et al. 1998).

Tyrosinase, the third marker used in this study, codes for an enzyme involved in the melanin pathway. Mutated forms of this enzyme result in an albino phenotype, due to nonfunctional pigment production (Giebel et al. 1991). Even though only one copy was previously described from each *Oryzias latipes* (Inagaki et al. 1994) and *Danio rerio* (Camp and Lardelli 2001), two copies are present in the pufferfish genome sequence.

For all three markers, an ancient duplication event in the fish lineage was inferred from previously performed phylogenetic analysis on *Danio rerio* and *Takifugu rubripes* sequences. We obtained sequences for basal actinopterygian species for these three markers and performed phylogenetic analyses based on the resulting amino acid alignments. From the concatenated datasets we inferred the most likely position of the duplication between the divergence of the Semionotiformes and the Osteoglossomorpha from the stem lineage of fish.

Materials and Methods

DNA Sources and Extraction

In this study, we included seven actinopterygian fish (*Acipenser baerii, Lepisosteus platyrhynchus*, *Gnathonemus petersi, Arapaima gigas*, *Elops hawaiensis, Oreochromis niloticus*), representing the major basal lineages as well as teleosts. DNA was extracted from muscle tissue stored at −80°C using the ATL extraction buffer (QIAGEN, Germany) and additional Proteinase K (final concentration, 1 mg/mL). After homogenization, DNA was purified by a standard phenol/chloroform procedure followed by ethanol precipitation (Sambrook et al. 1989).

PCR Amplification and Sequencing

Three molecular nuclear-encoded genes, *fzd8, sox11*, and *tyrosinase*, were chosen for this study since they are all present in duplicate in teleosts (fugu and zebrafish). The first two were chosen because they only consist of a single exon, facilitating PCR amplification from genomic DNA. The *tyrosinase* gene consists of five exons; the first one has a total length of 285 amino acid residues and was therefore targeted for phylogenetic analyses.

Degenerate primers were designed based on a nucleotide alignment including sequences from human, mouse, chicken, frog, zebrafish, and pufferfish. Different primer combinations amplified overlapping fragments of the selected genes (see Table 1 for universal fish primer sequences and Table 2 for species specific primers). PCR was performed in 50-μL reactions containing 1–1.5 units of REDTaq DNA polymerase (Sigma), 0.02 unit of *Pwo* DNA polymerase (Peqlab Biotechnology), 100 ng of genomic DNA, 20

Table 2. List of species-specific primers used in this study

Primer name	Organism	Sequence 5′ → 3′	Length	aa motif
Lepis.FZD8a.F880	*Lepisosteus*	G GAA GGA CTC TGG TCA GTG C	20	DGLWSV
Acip.FZD8.F880	*Acipenser*	G GAT GGA CTT TGG TCA GTG C	20	DGLWSV
Lepiso.FZD.F243 (aa)	*Lepisosteus*	CC ACT TTT GCC ACG GTT GCC	20	STFATVA
Lepiso.FZD.F270 (aa)	*Lepisosteus*	T TCA GTT GGC TAG ATC GTG	21	SVGYIVR
Gnatho.FZD.F190 (aa)	*Gnathonemus*	G GTG GAG GTG AAT GOT GAC C	20	VQVNGD
Gnatho.FZD.F245 (aa)	*Gnathonemus*	C GTC TCC ACC TTC GCC ACC	19	VSTFAT

Note. The primers designed for a genome walking approach in gar (*Lepisosteus*) and the elephant-nose fish (*Gnathonemus petersi*) are labeled with their position in an amino acid alignment.

pmol of each primer, 25 nmol of each dNTP, 50 nmol of additional $MgCl_2$, and the REDTaq PCR reaction buffer (onefold concentrated: 10 m*M* Tris–HCl, pH 8.3, 50 m*M* KCl, 1.1 m*M* $MgCl_2$, and 0.01% gelatin). Cycle conditions were adapted from a long-range PCR protocol (Barnes 1994), with an initial denaturation step at 94°C for 5 min, then 10 cycles at 94°C for 10 s, with annealing temperatures increasing by 0.5°C per cycle from 50 to 55°C but annealing time decreasing by 5 s per cycle from 90 to 40 s, and an extension step of 4 min at 68°C. Additional 20 cycles were performed at 94°C for 10 s, 55°C for 40 s, and 68°C for 4 min. The final extension was done at 68°C for 5 min. PCR products were purified either directly via spin columns (QIAGEN) or over gel (1% agarose) using the gel purification kit (QIAGEN). Sequencing was either performed directly using the corresponding PCR primers or cloned into the pCR2.1/TOPO vector (Invitrogen) and then sequenced using the M13 primers (forward and reverse).

DNA sequences of both strands were obtained using the Big-Dye Terminator cycle-sequencing ready reaction kit (Applied Biosystems Inc.) on an ABI-Hitachi 3100 capillary sequencer following the manufacturer's instructions.

Genome Walking

Genomic DNA was digested using blunt-end restriction enzymes, which were previously tested to produce fragments of usable size. The digests were purified by ethanol precipitation and double-stranded adapters consisting of GWA.mod.42 (5′ CGA CTC ACT ATA GGG CAC GCG TGG TCG ACG GCC CGG GCT GGT 3′) and GWA.short8.NH2 (5′ ACC AGC CCG CC-NH_2 3′) were ligated to the fragments. Two rounds of PCR were performed with two sequence-specific primers and two adapter-specific primers AP1.mod20 (5′-CA CTC ACT ATA GGG CAC GC-3′) and NAP2.mod18 (5′-GGC ACG CGT GGT CGA CGG-3′). This approach was applied for *Lepisosteus platyrhynchus* and *Gnathonemus petersi* to increase the length of the *fzd8* sequences.

Cloning of PCR Products

Some of the amplified PCR fragments produced ambiguous sequences indicating multiple amplified genes. These fragments were cloned with the TOPO-TA cloning kit (Invitrogen) to obtain possible paralogous sequences. From each transformation, 10 clones were sequenced. When multiple sequences of the expected gene were obtained from the clones, neighbor-joining analyses as implemented in MUST (Philippe 1993) were applied to confirm the sequence identity.

Phylogenetic Analyses

Nucleotide sequences were translated to amino acid sequences and aligned using ClustalX (Thompson et al. 1997). These alignments were conducted with the MUST package (Philippe 1993) and manually refined when necessary. Positions with gaps were eliminated, as were positions that could not be aligned unambiguously.

Maximum likelihood (ML) analyses were performed using TREE-PUZZLE 5.0 (Schmidt et al. 2002). The MEGA2.1 package (Kumar et al. 2001) was used for MP analyses (closest-neighbor interchange [CNI] on three levels), but also for NJ and ME analyses applying the gamma parameter estimated in the TREE-PUZZLE analysis. For the latter analyses, 1000 bootstrap replicates were performed. Also included in this study were ML methods based on Bayesian inference using MrBayes (Huelsenbeck and Ronquist 2001). We used the GTR (general time reversible) model in order to get the best estimates for every dataset because parameters are estimated for every possible amino acids substitution (Rodríguez et al. 1990).

Alternative topologies were compared with the ML tree applying two different likelihood-ratio tests: first, the Kishino–Hasegawa (KH; 1989) test and the Shimodaira–Hasegawa (SH; 1999) test (SH) as implemented in PAML (Yang 1997). The approximately unbiased (AU) test (Shimodaira 2002) as implemented in the CONSEL package was also applied, using the sidewise likelihood values (JTT model) estimated by PAML as starting point.

Results

DNA sequences for both copies of *fzd8* and *sox11* were available in GenBank from zebrafish, and duplicates for *fzd8* and *tyrosinase* from pufferfish were retrieved from the database of the Joint Genome Institute (JGI). Sequences for single copies of these genes were also previously described for rainbow trout, catfish, and medaka (see Table 3 for GenBank accession numbers). We determined *fzd8* sequences from *Acipenser baerii*, *Lepisosteus platyrhynchus*, *Gnathonemus petersi*, and *Oreochromis niloticus*; sequenced *sox11* genes from *A. baerii, L. platyrhynchus*, *G. petersi*, *Arapaima gigas*, *Elops hawaiensis, O. niloticus* and *Amphilophus citrinellum*; and determined new tyrosinase sequences from *A. baerii*, *L. platyrhynchus*, *G. petersi, E. hawaiensis*, and *O. niloticus*. We uncovered duplicates of *sox11* in the elephant-nose fish and of *tyrosinase* in Tilapia. For all other newly sequenced genes, only a single ortholog from each species was identified, although the PCR primers were designed to amplify both potential paralogs of each particular gene. PCR primers were tested for this property in the zebrafish, where the primers

Table 3. Taxa included in this study

Order	Family	Taxon	GenBank accession No.		
			fzd8	*sox11*	*tyr*
Actinopterygii					
Acipenseriformes	Acipenseridae	*Acipenser baerii*	AY333968	AY333969	AY333970
Semionotiformes	Lepisosteidae	*Lepisosteus platyrhynchus*	AY333980	AY333981	AY333982
Osteoglossomorpha (SD)					
Osteoglossiformes	Osteoglossidae	*Arapaima gigas*		AY333972	
	Mormyridae	*Gnathonemus petersi*	AY333976	AY333977 AY333978	AY333979
Elopomorpha (SD)					
Elopiformes	Elopidae	*Elops hawaiiensis*	AY333973	AY333974	AY333975
Euteleostei (SD)					
Cypriniformes	Cyprinidae	*Danio rerio*	AAD05435, AAD17520	NP_571411, NP_571412	AAN17339
Siluriformes	Ictaluridae	*Ictalurus punctatus*			AAF20161
Salmoniformes	Salmonidae	*Oncorhynchus mykiss*		BAA24575	
Beloniformes	Adrianichthyidae	*Oryzias latipes*			BAA06156
Tetraodontiformes	Tetraodontidae	*Takifugu rubripes*	JGI21332, JGI14550	JGI7177	JGI12109, JGI2193
Perciformes	Cichlidae	*Oreochromis niloticus*	AY333986	AY333983	AY333984 AY333985
		Amphilophus citrinellum		AY333971	
Sarcopterygii					
Class					
Amphibia	Pipidae	*Xenopus laevis*	AAC77361, AAC31121	BAA13006, BAA22779	AY333967
	Ranidae	*Rana nigromaculata*			BAA02077
Reptilia	Trionychidae	*Trionyx sinensis*			AAB25511
Aves	Gallidae	*Gallus gallus*		AB012237	P55024
	Phasianidae	*Coturnix japonicus*			BAB79631
Mammalia	Hominidae	*Homo sapiens*	NP_114072	P35716	AAB37227
	Muridae	*Mus musculus*	NP_032084	NP_033260	P11344
		Rattus norvegicus		NP_445801	
	Bovidae	*Bos taurus*			AAL38168
	Canidae	*Canis familiaris*			P54834

Note. SD, subdivision. All JGI numbers for *Takifugu rubripes* sequences refer to scaffold numbers of release v.1.0, October 26, 2001.

amplified both copies of the genes. As a further test of the methods employed, one fragment of *D. rerio* from the *fzd8* and one from the *sox11* genes were cloned and found that the primers are able to amplify both existing copies in this species. Cloning of single PCR fragments from other organisms (*E. hawaiensis*, *L. platyrhynchus*) did not provide additional genes, hence we conclude that those additional gene copies are likely not to be present in the genomes of the tested species.

Frizzled-8 (fzd8)

The alignment of this dataset consisted of nine sequences from ray-finned fish and four tetrapod outgroup sequences. The total length was 533 amino acid positions.

The phylogenetic analyses of the dataset (Fig. 2) strongly supported the monophyly of tetrapods and mammals and placed the two sequences of the tetraploid frog *Xenopus laevis* in a phylogenetic correct position. The position of *A. baerii* at the base of the actinopterygian cluster was recovered with good support by all methods applied, but within the Teleostei, there was no strong support for most of the internal nodes, due to the rather slow rate of evolution of these genes. The analysis showed that the two pufferfish genes are not the result of a pufferfish lineage specific duplication event. The duplication of the *fzd8* gene occurred at the latest in the common ancestor of pufferfish and zebrafish (indicated by an asterisk in Fig. 2), but firmer conclusions about the phylogenetic timing of the duplication of *fzd8* gene could not be drawn from the phylogenetic analyses of these gene sequence (Fig. 2).

We performed KH likelihood ratio tests to compare alternative topologies, but rearrangements among the teleost tree topologies were achieved with only minor changes of the likelihood value and, therefore, could not be rejected at the 5% significance level. However, a phylogenetic postion of *A. baerii* on a branch with the known duplicates from *D. rerio* and

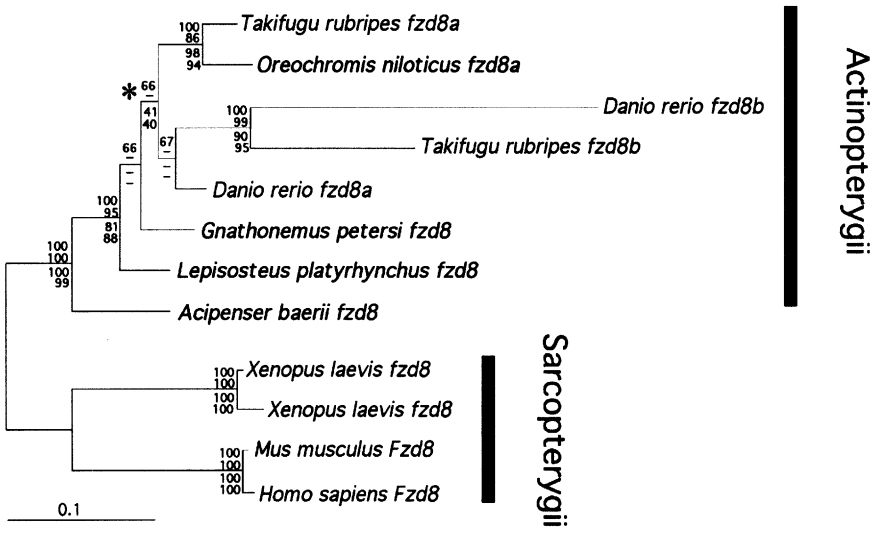

Fig. 2. Maximum likelihood tree of the *fsd8* dataset (12 sequences, 533 positions) as obtained by Bayesian inference. Sequences in boldface were obtained in this study. Numbers above branches indicate posterior probabilities (MrBayes; upper value of quartet), Quartet Puzzling support values (TREE-PUZZLE; second value of quartet), and bootstrap values from neighbor joining (third value of quartet) and maximum parsimony (both MEGA2.1; lowest value of quartet). Only values above 40% are shown. The asterisk indicates the inferred phylogenetic timing of the fish-specific genome duplication.

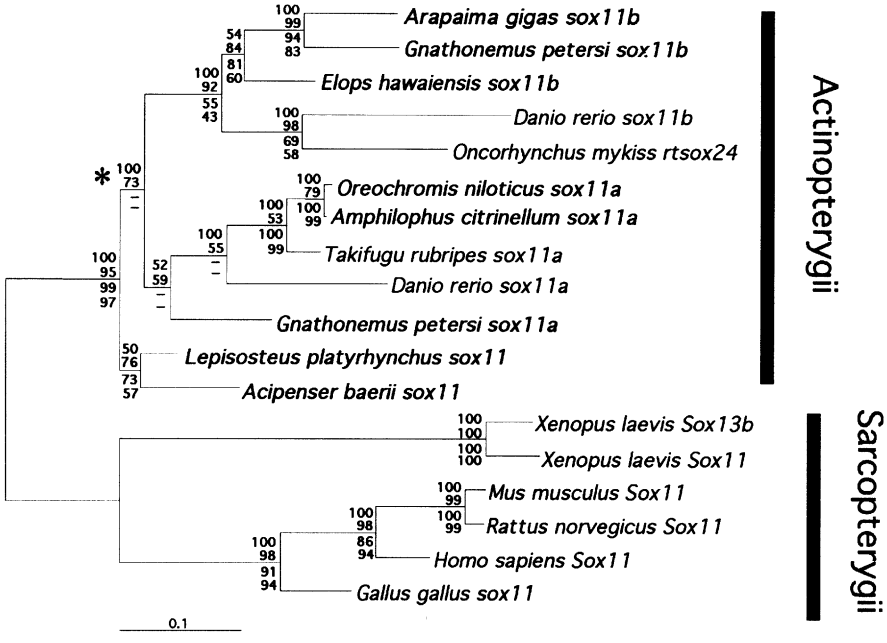

Fig. 3. Maximum likelihood tree of the *sox11* dataset (18 sequences, 319 positions) as obtained by Bayesian inference. Sequences in boldface were obtained in this study. Numbers above branches indicate posterior probabilities (MrBayes; upper value of quartet), Quartet Puzzling support values (TREE-PUZZLE; second value of quartet), and bootstrap values from neighbor joining (third value of quartet) and maximum parsimony (both MEGA2.1; lowest value of quartet). Only values above 40% are shown. The asterisk indicates the inferred phylogenetic timing of the fish-specific genome duplication.

T. rubripes was significantly rejected by the KH test (data not shown), indicating that the duplication of *fzd8* genes occurred after the sturgeons and their relatives branched off the fish stem lineage.

Sox11

An amino acid alignment consisting of 18 sequences (12 actinopterygian and 6 tetrapod outgroup sequences) with 319 positions was created. Phylogenetic analyses of the data (Fig. 3) found good support for the monophyly of mammals, amniotes, and tetrapods. Within the highly supported monophyletic actinopterygian group, some of the branches remain weakly supported. The duplicates in *G. petersi* and *D. rerio* are clearly not sister sequences (differing by 10%–30% sequence difference based on the amino acid sequence; data not shown) and, therefore, cannot be the result of recent independent gene duplications within these lineages, but must be of rather ancient origin, the likely position of which is indicated by an asterisk (Fig. 3). This result was also confirmed by a significant KH likelihood ratio testing the different topologies (i.e., placing the two *sox11* genes from *G. petersi* as sister genes; data not shown).

Tyrosinase *Exon 1*

An alignment of 20 sequences (11 actinopterygian and 9 tetrapod outgroup sequences) with 238 amino acid positions was created. The analyses of the *tyrosinase* dataset (Fig. 4) clearly supported most of the

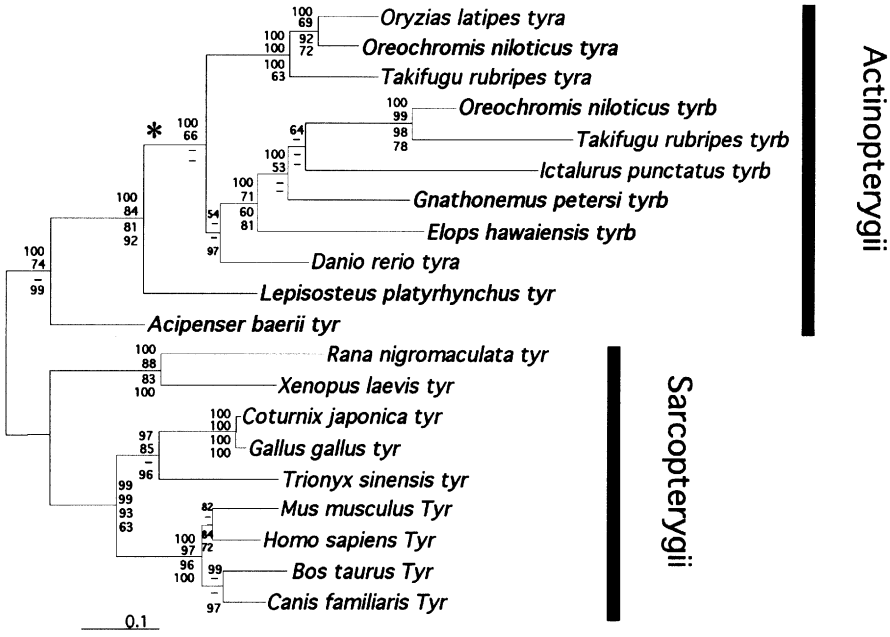

Fig. 4. Maximum likelihood tree of the *tyrosinase* (exon 1) dataset (20 sequences, 238 positions) as obtained by Bayesian inference. Sequences in boldface were obtained in this study. Numbers above branches indicate posterior probabilities (MrBayes; upper value of quartet), Quartet Puzzling support values (TREE-PUZZLE; second value of quartet), and bootstrap values from neighbor joining (third value of quartet) and maximum parsimony (both MEGA2.1; lowest value of quartet). Only values above 40% are shown. The asterisk indicates the inferred phylogenetic timing of the fish-specific genome duplication.

main tetrapod clades such as mammals, amniotes, and amphibians. Some phylogenetic methods, however, did not recover the monophyly of actinopterygians or reptiles and the expected relationships within the mammals. Within the actinopterygian lineage some nodes were only weakly supported. A sister-group relationship of the *I. punctatus* and the *D. rerio* sequences (both Ostariophysii) was rejected by the KH test, implying that their genes belong to different paralogy groups and, therefore, indicating independent gene losses on different evolutionary lineages from an ancient duplication (indicated by an asterisk in Fig. 4). In order to distinguish between the two discovered paralogs, we refer to the duplicates as *tyra* and *tyrb* from here on.

Concatenation of Datasets

Improved phylogenetic resolution with a combined, larger dataset is expected (Lecointre et al. 1994). The three datasets were concatenated, based on the assumption that the observed duplications for our markers all correspond to the same event, i.e., are all caused by a fish-specific genome duplication. If the gene duplicates were not produced by a single whole-genome duplication event, but through independent, smaller genomic events or tandem duplications at different times during the evolution of fish, one might expect an increase of noise and loss of phylogenetic signal.

Due to limited species availability for the different markers, it was sometimes necessary to combine sequences from different species that belong to the same larger fish taxon. For example, *O. mykiss rtSox24* (Protacanthopterygii) was combined with *T. rubripes tyrb* to create a combined sequence that was representative of the subdivision Euteleostei. Likewise, *D. rerio sox11b* was fused with the *I. punctatus tyrb* sequence to represent the suborder Ostariophysii.

A dataset of *sox11* and *tyrosinase* genes consisting of 14 sequences (9 ingroup and 5 outgroup) and a total length of 557 amino acid positions was analyzed with different phylogenetic methods (Fig. 5). The monophyly of tetrapods and the relationships among them are clearly resolved, as well as the basal position of *A. baerii* within the actinopterygians. A separation of the lineage leading to *L. platyrhynchus* from the evolutionary lineage leading to more modern fish—probably before the duplication event (indicated by an asterisk in Fig. 5)—is recovered by all methods applied. Among the orders that presumably originated after the fish-specific genome duplication, the phylogenetic resolution is weaker, especially in neighbor joining and maximum parsimony analyses.

Likelihood ratio tests (KH and SH tests) were applied to test alternative topologies that would be explained with different phylogenetic origins of duplicates genes (Table 4). Gene tree topologies with the gar being a part of the duplication event were significantly rejected by all three likelihood ratio tests, the KH, the SH, and the AU test. Hence, as suggested by the analyses of individual duplicates, the origin of the paralogs (due to a common genome-wide duplication) was estimated to have occurred after the gar lineage diverged from the fish stem lineage. Changing the position of the *E. hawaiensis* and *G. petersi* clade to the other branch of the duplication or outside the duplication was significantly rejected by the KH and the AU, but not by the SH, test.

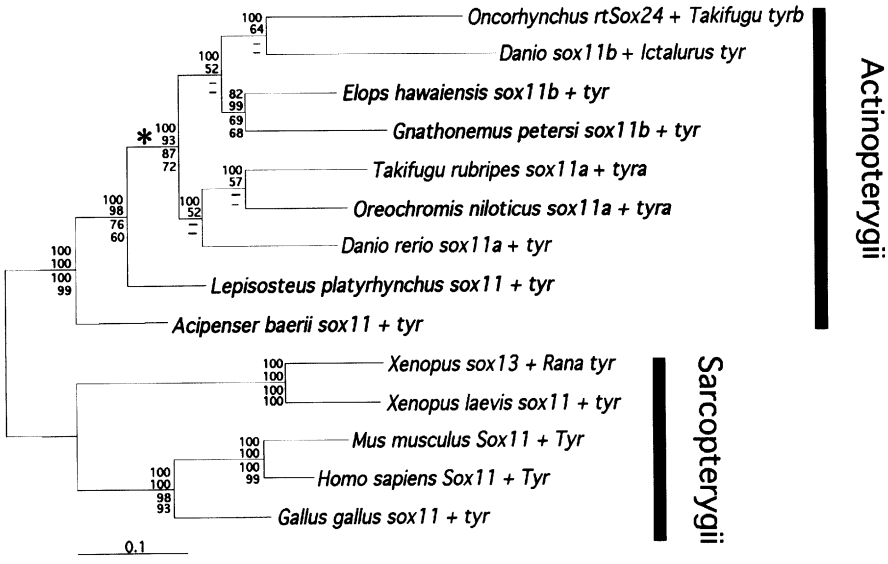

Fig. 5. Maximum likelihood tree of the *sox11* and *tyrosinase* (exon 1) dataset (14 sequences, 557 positions) as obtained by Bayesian inference. Sequences in boldface were obtained in this study. Numbers above branches indicate posterior probabilities (MrBayes; upper value of quartet), Quartet Puzzling support values (TREE-PUZZLE; second value of quartet), and bootstrap values from neighbor joining (third value of quartet) and maximum parsimony (both MEGA2.1; lowest value of quartet). Only values above 40% are shown. The asterisk indicates the inferred phylogenetic timing of the fish-specific genome duplication.

Table 4. Comparison of the likelihood values of different topologies within the Actinopterygii, applying the Kishino–Hasegawa test (KH) and the Shimodaira–Hasegawa-test (SH) based on the sox11-tyrosinase dataset

Topology	Li	ΔLi	±SE	*P*KH	*P*SH	*P*AU
1. (A(L(((G + E) (O + I)) (D(C + T)))))	−6782.639	0.000	0.000	−1.000	−1.000	0.839
2. (A(L(((G(E(O + I))) (D(C + T)))))	−6785.527	−2.887	3.137	0.179	0.685	0.228
3. (A((L((G + E) (O + 1))) (D(C + T))))	−6812.590	−29.951	11.394	**0.004**	**0.046**	**0.002**
4. (A(((G + E) (O + I)) (L(D(C + T)))))	−6812.568	−29.929	11.425	**0.004**	**0.047**	**0.001**
5. (A(L((G + E) ((O + I) (D(C + T))))))	−6809.472	−26.833	12.003	**0.013**	0.073	**0.006**
6. (A(L((O + I) ((G + E) (D(C + T))))))	−6809.938	−27.299	11.850	**0.011**	0.068	**3.004**

Note. The first topology is the maximum likelihood tree. Abbreviations: *Acipenser* (A), *Lepisosteus* (L), *Gnathonemus* (G), *Elops* (E), *Oncorhynchus* + *Takifugu* (O), *Danio* + *Ictalurus* (I), *Danio* (D), *Oreochromis* (C), *Takifugu* (T), likelihood (Li), difference of likelihood (ΔLi), *P* value for Kishino–Hasegawa (*P*KH), *P* value for Shimodaira–Hasegawa (*P*SH), *P* value for approximately unbiased test (*P*AU). Values in boldface indicate significance at the 5% level.

Concatenation of sox11, tyrosinase *Exon1, and* fzd8

Based on the previously mentioned datasets, one concatenated alignment for all three genes could be formed containing 12 sequences (8 ingroup and 4 outgroup) and is made up of 1090 amino acid positions. All methods applied reconstructed the same topology (Fig. 6). The tree is clearly resolved with high support for the tetrapod outgroup relationships, monophyly of actinopterygians, and the sturgeons (*A. baerii*) and gars (*L. platyrhynchus*) splitting off from the fish stem lineage before the inferred fish-specific genome duplication event.

Comparing different topologies with likelihood ratio tests, the KH and AU tests significantly ruled out the possibility of gars being originating after the duplication event. Changing the position of the osteoglossomorphs (*G. petersi*) to the branch with the paralogous sequences was also rejected by KH and AU tests. For the *D. rerio* sequence, a change in position to the other paralogous group or prior to the presumed duplication (indicated by an asterisk in Fig. 6) was significant rejected. A monophyletic grouping of sturgeons and gars (*A. baerii* and *L. platyhynchus*), as recently proposed based on mitochondrial DNA data of Inoue et al. (2003), was rejected by both the KH and the AU tests. The SH test failed to reject any of the tested topologies.

Discussion

Until now, for most duplicated genes, only sequences from rather young, derived fish lineages (Euteleostei), i.e., the zebrafish and pufferfish, were available. The goal of this study was to identify which of the early-branching lineages of actinopterygian fish diverged from the fish stem lineage before the presumed fish-specific genome duplication occurred. To this end, we amplified orthologous genes from basal fish lineages, which are known to be duplicated in teleosts. Our analyses support the hypothesis that a fish-specific genome duplication event that occurred in the fish stem lineage after the separation of gars (e.g., *Lepisosteus platyrhynchus*) but before the origin of the Osteoglossomorpha (Figs. 5 and 6).

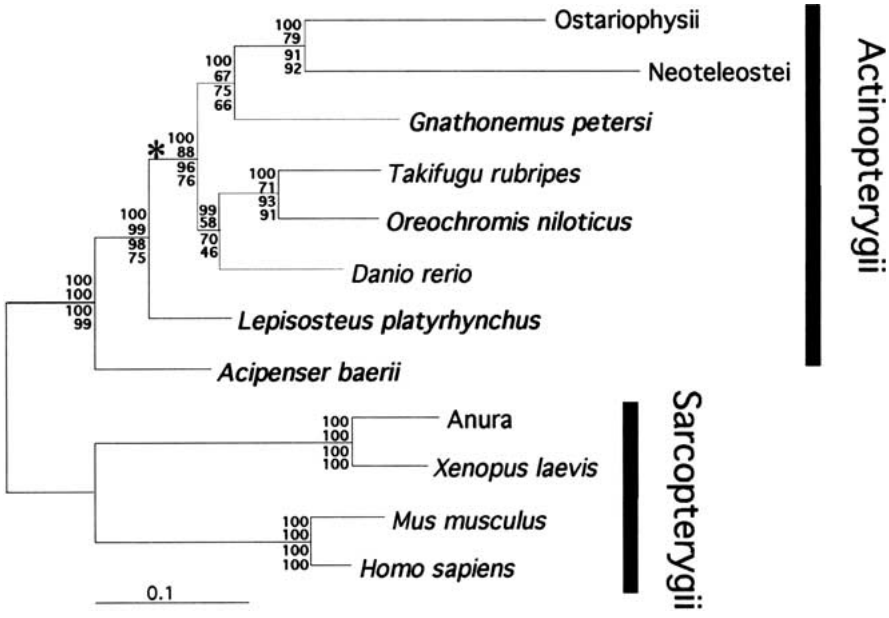

Fig. 6. Maximum likelihood tree of concatenated dataset of *sox11*, *tyrosinase* (exon 1), and *fzd8* (12 sequences, 1090 positions) as obtained by Bayesian inference. Sequences in boldface were obtained in this study. Numbers above branches indicate posterior probabilities (MrBayes; upper value of quartet), Quartet Puzzling support values (TREE-PUZZLE; second and value of quartet), and bootstrap values from neighbor joining (third value of quartet) and maximum parsimony (both MEGA2.1; lowest value of quartet). Only values above 40% are shown. The asterisk indicates the inferred phylogenetic timing of the fish-specific genome duplication.

If our interpretation of the data were correct, we would expect, under the simplest scenario (no independent gene loss), to find, for each single-copy gene in sturgeons, bichir, and the gars, two paralogous copies in all fish belonging to the division Teleostei (Osteoglossomorpha, Elopomorpha, Euteleostei). However, in only two instances (*Gnathonemus petersi sox11a/b* and *Oreochromis niloticus tyra/b*) did we find two genes in these derived fish. We cloned candidate fragments from *L. platyrhynchus, E. hawaiensis*, and *G. petersi* and sequenced 10 clones each but found no additional copies. The expected sequence identity for the paralogs was 70%–90% (at the amino acid level), based on data for other ancient, duplicated genes, but we found only minor allelic variation (data not shown). The lack of these copies can be interpreted as indicating either a strong amplification bias or, more likely, that one paralog was secondarily lost during evolution.

Although gene loss is a frequent event, 20%–50% of paralogous genes are retained for longer evolutionary time spans after a genome duplication event (Postlethwait et al. 2000; Lynch and Force 2000a). Sidow (1996) argued that if one of those copies experiences a mutation that enables it to perform a single unique function in an ocean of redundancy (neofunctionalization), this selective advantage would be sufficient to retain this gene copy and to prevent degenerative substitutions and prevent this functional gene copy from becoming a pseudogene. Another possibility is a division of the original function between the duplicates (subfunctionalization) (Force et al. 1999); also, this process would tend to retain both gene copies after a duplication event.

The evolutionary rates of two paralogs often differ enormously; usually one of the paralogs evolves considerably faster than the other one (Van de Peer et al. 2001). This phenomenon can lead to problems in phylogenetic reconstruction, but also reduces the efficiency of degenerate PCR primers and therefore might result in a biased amplification of only one copy. Hence, we cannot rule out with certainty that some of the presumed duplicated genes could not be detected by our methodology.

The phylogenetic reconstruction of the expected topology that one would expect based on a genome duplication in the fish lineage is problematic for many duplicated genes, probably due to different evolutionary rates between sequences (Taylor et al. 2001a) and saturation effects (Van de Peer et al. 2002a). Another way of getting better resolution of phylogenetic relationships is to add more data, either by more positions in the dataset (Lecointre et al. 1994) or by adding more sequences/species, since larger concatenated datasets often improve the resolution of more difficult phylogenetic relationships (Miya and Nishida 2000; Nei et al. 2001; Zwickl and Hillis 2002).

For both *sox11* and *fzd8*, the reconstruction of the expected duplication topology was problematic, with the original dataset consisting of only four teleost sequences and the tetrapod outgroup. Adding more sequences increased the support for the duplication event within the *sox11* gene tree (Fig. 3). This was not the case for *fzd8*, where the *Danio rerio fzd8a* sequence remains positioned at the base of the *fzd8b* branch.

For our data, the bootstrap support clearly increases with the length of the sequences, although likelihood ratio test failed to reject alternative topologies for the concatenated dataset with all three markers. We applied different tests, the KH test (Kishino and Hasegawa 1989) and the SH test (Shi-

modaira and Hasegawa 1999). While the KH test still rejected some alternative topology for the larger datasets, the SH test failed to reach significant support. Recently, the SH test was found to be too conservative in some cases, which means that it failed to reject a wrong hypothesis (Strimmer and Rambaut 2002). The AU test (Shimodaira 2002) was significant for most topologies we tested for the sox11–tyrosinase concatemer (Fig. 5) and, also, for the triple dataset (Fig. 6). Based on these likelihood ratio tests, we are confident that the genome duplication event took place after the split of the gar lineage.

Phylogeny of Actinopterygian Fish

Mitochondrial data were able to resolve many relationships among the teleosts (Inoue et al. 2001; Miya et al. 2003) and, also, gave a well-supported tree of the more basal splits (Inoue et al. 2003; Noack et al. 1996) which were previously uncertain, e.g., the question of the closest relative of the Teleostei (Arratia 2001). Even though mitochondrial data are more easily available than nuclear genes, previous papers have demonstrated their limits for phylogenetic reconstruction, particularly for the estimation of relationships that date back more than about 300–350 mya (Zardoya and Meyer 2001; Meyer and Zardoya 2003).

In contrast to mitochondrial data, our nuclear markers weakly support the Elopomorpha and the Osteoglossomorpha forming a monophyletic group, a topology previously suggested in a phylogenetic analysis of 28S-rRNA (Le et al. 1993). Neither of these nuclear markers, though, clearly rejected the hypothesis of the Osteoglossomorpha being at the basal position of the Teleostei, which is clearly supported by the mtDNA (Inoue et al. 2001). The sister-group relationship of Acipenseriformes and Semionotiformes (Inoue et al. 2003) was rejected by our data, which suggests the more traditional view of the Acipenseriformes lineage branching off earlier.

For an estimation of age of the duplication, the dates of origins of the major lineages of fish based on fossils are considered. However, paleontological and molecular estimates often differ widely (e.g., Kumar and Hedges 1998; Hedges and Kumar 2003; Meyer and Zardoya 2003). The paleontological data, of course, can only provide minimum ages, and those are usually much younger than the estimates based on molecular data. Fossil finds date the age of the Semionotiformes to between 245 and 286 mya (Permian) (Wiley and Schultze 1984), while molecular estimates for the Amiiformes, which are approximately of the same age as Semionotiformes, hint at a separation from the Teleostei stem lineage about 367–404 mya (Kumazawa et al. 1999). Molecular data suggest an age of 335 mya for the Osteoglossomorpha (Kumazawa and Nishida 2000). From these data, the fish-specific genome duplication can be dated to between 335 and 404 mya. These findings are in agreement with recent analyses of the complete pufferfish genome, which showed an increased amount of duplicated genes that originated 320 ± 67 mya (Vandepoele et al. 2004).

The Fish-Specific Genome Duplication and the Radiation of Fish

Previous studies suggested that the fish-specific genome duplication provided a genomic mechanism and impetus for the explosive radiation of the almost 25,000 species of teleost fish (Amores et al. 1998; Wittbrodt et al. 1998; Meyer and Schartl 1999). The small internodes between the duplication event and the separation of the Osteoglossomorpha and the Elopomorpha might be an indication of fast lineage origination and increased rates of speciation following this event. Orders of fish, which originated after the fish-specific genome duplication, are strikingly more species-rich than the more basal actinopterygians, again lending support for a connection between genomic gene content and presumably resulting in increased complexity of gene networks and species diversification and increased phenotypic complexity.

Gene silencing and subsequent loss can happen within a short time after a gene duplication event (Li 1980; Lynch and Conery 2000, 2003; Lynch 2002), and divergent resolution has been proposed as a mechanism leading to an increase in the rate of speciation (Taylor et al. 2001b). Divergent resolution (Lynch and Force 2000b; Taylor et al. 2001c), the loss of different paralogs in different populations, might lead to genetic isolation and speciation in populations that retained different sets of paralogs.

Conclusions

From our data from three nuclear molecular markers, we propose that the fish-specific genome duplication event took place between the split of the Semionotiformes (*Lepisosteus platyrhynchus*) from the fish stem lineage and the origin of the Osteoglossomorpha (335–404 mya). The fish-specific genome duplication might be causally related to an increase in species and morphological diversity. The phylogenetic timing of the fish-specific duplication event that is supported by our data is between the origin of nonteleostean, actinopterygian groups, consisting of 44 species in 5 families, and the division Teleostei, which contains 23,637 species in 425 families (Nelson 1994). The subdivisions Osteoglosso-

morpha (217 species) and Elopomorpha (37 species), as the first lineage to diverge from the fish stem lineage after the presumed genome duplication event, show an elevated number of species compared to more basal actinopterygians (44 species in five families). The difference in numbers of species in the lineages that separated from the fish stem lineage before the duplication and lineages which originated after the genome duplication is striking and might indeed indicate a causal link of this genome event and the realised genetic potential in terms of speciation. Future work will be required to further investigate how regulatory evolution and evolution by duplication work independently or jointly to facilitate evolutionary diversification.

Acknowledgments. We thank members of the Meyer lab, in particular, Yves Van de Peer and Walter Salzburger, for discussion. This work was supported by the University of Konstanz and by grants from the Deutsche Forschungsgemeinschaft and the Fond der Chemischen Industrie to A.M.

References

Acampora D, D'Esposito M, Faiella A, Pannese M, Migliaccio E, Morelli F, Stornaiuolo A, Nigro V, Simeone A, Boncinelli E (1989) The human HOX gene family. Nucleic Acids Res 17:10385–10402

Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL, Westerfield M, Ekker M, Postlethwait JH (1998) Zebrafish hox clusters and vertebrate genome evolution. Science 282:1711–1714

Amores A, Suzuki T, Yan Y, Pomeroy J, Singer A, Amemiya C, Postlethwait JH (2004) Developmental roles of pufferfish Hox clusters and genome evolution in ray-fin fish. Genome Res 14:1–10

Aparicio S, Chapman J, Stupka E, Putnam N, Chia Jm, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MDS, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoef F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJK, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D, Brenner S (2002) Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. Science 297:1301–1310

Arratia G (2001) The sister-group of Teleostei: Consensus and disagreements. J Vert Paleontol 21:767–772

Barnes WM (1994) PCR amplification of up to 35-kb DNA with high fidelity and high yield from λ bacteriophage templates. Proc Natl Acad Sci USA 91:2216–2220

Barbazuk BB, Korf I, Kadavi C, Heyen J, Tate S, Wun E, Bedell JA, McPherson JD, Johnson SL (2000) The syntenic relationship of the zebrafish and human genomes. Genome Res 10:1351–1358

Bowles J, Schepers G, Koopman P (2000) Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. Dev Biol 227:239–255

Camp E, Lardelli M (2001) Tyrosinase gene expression in zebrafish embryos. Dev Genes Evol 211:150–153

Chiang EFL, Pai C-I, Wyatt M, Yan Y, Postlethwait J, Chung B (2001) Two Sox9 Genes on Duplicated Zebrafish Chromosomes: Expression of similar transcription activators in distinct sites. Dev Biol 231:149–163

Eldredge N, Stanley SM (1984) Living fossils. Springer-Verlag, New York

Force A, Lynch M, Pickett FB, Amores A, Yan Y, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545

Garcia-Fernandez J, Holland PW (1994) Archetypal organization of the amphioxus Hox gene cluster. Nature 370:563–566

Gates MA, Kim L, Egan ES, Cardozo T, Sirotkin HI, Dougan ST, Lashkari D, Abagyan R, Schier AF, Talbot WS (1999) A genetic linkage map for zebrafish: Comparative analysis and localization of genes and expressed sequences. Genome Res 9:334–347

Giebel LB, Tripathi RK, Strunk KM, Hanifin JM, Jackson CE, King RA, Spritz RA (1991) Tyrosinase gene mutations associated with type IB (''yellow'') oculocutaneous albinism. Am J Hum Genet 48:1159–1167

Graham A, Papalopulu N, Krumlauf R (1989) The murine and Drosophila homeobox gene complexes have common features of organization and expression. Cell 57:367–378

Grande L, Bemis WE (1996) Interrelationships of Acipenseriformes, with comments on ''Chondrostei.'' In: Stiassny MLJ, Parenti LR, Johnson GD (eds) Interrelationships of fishes. Academic Press, London, pp 85–115

Harvey RP, Tabin CJ, Melton DA (1986) Embryonic expression and nuclear localization of Xenopus homeobox (Xhox) gene products. EMBO J 5:1237–1244

Hedges SB, Kumar S (2003) Genomic clocks and evolutionary timescales. Trends Genet 19:200–206

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755

Inagaki H, Bessho Y, Koga A, Hori H (1994) Expression of the tyrosinase-encoding gene in a colorless melanophore mutant of the medaka fish, Oryzias latipes. Gene 15:319–324

Inoue JG, Miya M, Tsukamoto K, Nishida M (2001) A mitogenomic perspective on the basal teleostean phylogeny: Resolving higher-level relationships with longer DNA sequences. Mol Phylogenet Evol 20:275–285

Inoue JG, Miya M, Tsukamoto K, Nishida M (2003) Basal actinopterygian relationships: A mitogenomic perspective on the phylogeny of the ''ancient fish.'' Mol Phylogenet Evol 26:110–120

Kao H, Lee S-C (2002) Phosphoglucose isomerases of hagfish, zebrafish, gray mullet, toad, and snake, with reference to the evolution of the genes in vertebrates. Mol Biol Evol 19:367–374

Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. J Mol Evol 29:170–179

Kumar S, Hedges SB (1998) A molecular timescale for vertebrate evolution. Nature 392:917–920

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17:1244–1245

Kumazawa Y, Nishida M (2000) Molecular phylogeny of osteoglossoids: A new model for Gondwanian origin and plate tectonic transportation of the Asian arowana. Mol Biol Evol 17:1869–1878

Kumazawa Y, Yamaguchi M, Nishida M (1999) Mitochondrial molecular clocks and the origin of euteleostean biodiversity: Familial radiation of Perciforms may have predated the Cretaceous/Tertiary boundary. In: Kato M (ed) The biology of biodiversity. Springer-Verlag, Hong Kong, pp 35–52

Le HL, Lecointre G, Perasso R (1993) A 28S rRNA-based phylogeny of the gnathostomes: First steps in the analysis of conflict and congruence with morphologically based cladograms. Mol Phylogenet Evol 2:31–51

Lecointre G, Philippe H, Van Le HL, Le Guyader H (1994) How many nucleotides are required to resolve a phylogenetic prob-

lem? The use of a new statistical method applicable to available sequences. Mol Phylogenet Evol 3:292–309

Li WH (1980) Rate of gene silencing at duplicate loci: A theoretical study and interpretation of data from tetraploid fishes. Genetics 95:237–258

Lister JA, Close J, Raible DW (2001) Duplicate *mitf* genes in zebrafish: Complementary expression and conservation of melanogenic potential. Dev Biol 237:333–344

Lynch M (2002) Gene duplication and evolution. Science 297:945–947

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

Lynch M, Force A (2000b) The origin of interspecific genomic incompatibility via gene duplication. Am Nat 156:590–605

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155

Lynch M, Conery JS (2003) The evolutionary demography of duplicate genes. J Struct Funct Genomics 3:35–44

Málaga-Trillo E, Meyer A (2001) Genome duplications and accelerated evolution of Hox genes and cluster architecture in teleost fishes. Am Zool 41:676–686

Martinez-Barbera JP, Toresson H, Da Rocha S, Krauss S (1997) Cloning and expression of three members of the zebrafish Bmp family: Bmp2a, Bmp2b and Bmp4. Gene 198:53–59

Merritt TJS, Quattro JM (2001) Evidence for a period of directional selection following gene duplication in a neurally expressed locus of triosephosphate isomerase. Genetics 159:689–697

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

Meyer A, Van de Peer Y (eds) (2003) Genome evolution: Gene and genome duplications and the origin of novel gene functions. Kluwer Academic, Dordrecht

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

Miya M, Nishida M (2000) Use of mitogenomic information in teleostean molecular phylogenetics: A tree-based exploration under the maximum-parsimony optimality criterion. Mol Phylogenet Evol 17:437–455

Miya M, Takeshima H, Endo H, Ishiguro NB, Inoue JG, Mukai T, Satoh TP, Yamaguchi M, Kawaguchi A, Mabuchi K, Shirai SM, Nishida M (2003) Major patterns of higher teleostean phylogenies: A new perspective based on 100 complete mitochondrial DNA sequences. Mol Phylogenet Evol 26:121–138

Naruse K, Fukamachi S, Mitani H, Kondo M, Matsuoka T, Kondo S, Hanamura N, Morita Y, Hasegawa K, Nishigaki R, Shimada A, Wada H, Kusakabe T, Suzuki N, Kinoshita M, Kanamori A, Terado T, Kimura H, Nonaka M, Shima A (2000) A detailed linkage map of medaka, Oryzias latipes: Comparative genomics and genome evolution. Genetics 154:1773–1784

Nei M, Xu P, Glazko G (2001) Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. Proc Natl Acad Sci USA 98:2497–2502

Nelson J (1994) Fishes of the world. Wiley, New York

Noack K, Zardoya R, Meyer A (1996) The complete mitochondrial DNA sequence of the bichir (Polypterus ornatipinnis), a basal ray-finned fish: Ancient establishment of the consensus vertebrate gene order. Genetics 144:1165–1180

Normark BB, McCune AR, Harrison RG (1991) Phylogenetic relationships of neopterygian fishes, inferred from mitochondrial DNA sequences. Mol Biol Evol 8:819–834

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

Ohno S (1999) Gene duplication and the uniqueness of vertebrate genomes circa 1970–1999. Semin Cell Dev Biol 10:517–522

Patterson C (1973) Interrelationships of holosteans. In: Greenwood H, Miles RS, Patterson C (eds) Interrelationships of fishes. Zool J Linn Soc Lond (Suppl 1): 233–305

Philippe H (1993) MUST, a computer package of Management Utilities for Sequences and Trees. Nucleic Acids Res 21:5264–5272

Postlethwait JH, Woods IG, Ngo-Hazelett P, Yan Y-L, Kelly PD, Chu F, Huang H, Hill-Force A, Talbot WS (2000) Zebrafish comparative genomics and the origins of vertebrate chromosomes. Genome Res 10:1890–1902

Rimini R, Beltrame M, Argenton F, Szymczak D, Cotelli F, Bianchi ME (1999) Expression patterns of zebrafish *sox11*A, *sox11*B and *sox21*. Mech Dev 89:167–171

Risinger C, Salaneck E, Soderberg C, Gates M, Postlethwait JH, Larhammar D (1998) Cloning of two loci for synapse protein Snap25 in zebrafish: comparison of paralogous linkage groups suggests loss of one locus in the mammalian lineage. J Neurosci Res 54:563–573

Robinson-Rechavi M, Marchand O, Escriva H, Laudet V (2001) An ancestral whole-genome duplication may not have been responsible for the abundance of duplicated fish genes. Curr Biol 11:R458–R459

Rodríguez FJ, Oliver A, Marìn A, Medina JR (1990) The general stochastic model of nucleotide substitution. J Theor Biol 142:485–501

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning—A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18:502–504

Schultze HP, Wiley, EO (1984) The neopterygian *Amia* as a living fossil. In: Eldredge N, Stanley SM (eds) Living fossils. Springer-Verlag, New York, pp 153–159

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

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

Sidow A (1996) Gen(om)e duplications in the evolution of early vertebrates. Curr Opin Genet Dev 6:715–722

Spring J (1997) Vertebrate evolution by interspecific hybridisation—Are we polyploid? FEBS Lett 400:2–8

Strimmer K, Rambaut A (2002) Inferring confidence sets of possibly misspecified gene trees. Proc R Soc Lond B Biol Sci 269:137–142

Taylor JS, Van de Peer Y, Braasch 1, Meyer A (2001a) Comparative genomics provides evidence for an ancient genome duplication event in fish. Philos Trans R Soc Lond B Biol Sci 356:1661–1679

Taylor JS, Van de Peer Y, Meyer A (2001b) Revisiting a recent test of the ancient fish-specific genome duplication hypothesis. Curr Biol 11:R1005–R1007

Taylor JS, Van de Peer Y, Meyer A (2001c) Genome duplication, divergent resolution and speciation. Trends Genet 17:299–301

Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y (2003) Genome duplication, a trait shared by 22,000 species of ray-finned fish. Genome Res 13:382–390

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Van de Peer Y, Taylor JS, Braasch I, Meyer A (2001) The ghost of selection past: Rates of evolution and functional divergence of anciently duplicated genes. J Mol Evol 53:436–446

Van de Peer Y, Frickey T, Taylor JS, Meyer A (2002a) Dealing with saturation at the amino acid level: a case study based on anciently duplicated zebrafish genes. Gene 295:205–211

Van de Peer Y, Taylor JS, Jayabalan J, Meyer A (2002b) Wanda: A data base of duplicated fish genes. Nucleic Acids Res 30:109–112

Van de Poele K, De Vos W, Taylor JS, Meyer A, Van de Peer Y (2004) Major events in the genome evolution of vertebrates: Paranome age and size differs considerably between ray-finned fishes and land vertebrates. Proc Natl Acad Sci USA 101:1638–1643

Wang Y, Macke JP, Abella BS, Andreasson K, Worley P, Gilbert DJ, Copeland NG, Jenkins NA, Nathans J (1996) A large family of putative transmembrane receptors homologous to the product of the *Drosophila* tissue polarity Gene *frizzled*. J Biol Chem 271:4468–4476

Wiley EO, Schultze HP (1984) Family Lepisosteidae (gars) as living fossils. In: Eldredge N, Stanley SM (eds) Living fossils. Springer-Verlag, New York, pp 160–165

Wittbrodt J, Meyer A, Schartl M (1998) More genes in fish? BioEssays 20:511–515

Yang Z (1997) PAML: Program package for phylogenetic analysis by maximum likelihood. CABIOS 13:555–556

Zardoya R, Meyer A (2001) Vertebrate phylogeny: limits of inference of mitochondrial genome and nuclear rRNA sequence data due to an adverse phylogenetic signal/homoplasy ratio. In: Ahlberg P (ed) Major events in early vertebrate evolution: Palaeontology phylogeny and development. Taylor and Francis, London, pp 135–155

Zwickl DJ, Hillis DM (2002) Increased taxon sampling greatly reduces phylogenetic error. Syst Biol 51:588–598