DEVELOPMENTAL DYNAMICS 228:480–489, 2003

Genome Duplication, Subfunction Partitioning, and Lineage Divergence: *Sox9* in Stickleback and Zebrafish

William A. Cresko, Yi-Lin Yan, David A. Baltrus, Angel Amores, Amy Singer, Adriana Rodríguez-Marí, and John H. Postlethwait*

Teleosts are the most species-rich group of vertebrates, and a genome duplication (tetraploidization) event in ray-fin fish appears to have preceded this remarkable explosion of biodiversity. What is the relationship of the ray-fin genome duplication to the teleost radiation? Genome duplication may have facilitated lineage divergence by partitioning different ancestral gene subfunctions among co-orthologs of tetrapod genes in different teleost lineages. To test this hypothesis, we investigated gene expression patterns for *Sox9* gene duplicates in stickleback and zebrafish, teleosts whose lineages diverged early in Euteleost evolution. Most expression domains appear to have been partitioned between *Sox9a* and *Sox9b* before the divergence of stickleback and zebrafish lineages, but some ancestral expression domains were distributed differentially in each lineage. We conclude that some gene subfunctions, as represented by lineage-specific expression domains, may have assorted differently in separate lineages and that these may have contributed to lineage diversification during teleost evolution. *Developmental Dynamics 228:480–489, 2003.* © 2003 Wiley-Liss, Inc.

Key words: genome duplication; subfunctionalization; transcription factor; chondrogenesis; sex determination; macroevolution; tetraploidization

Received 28 July 2003; Accepted 4 August 2003

INTRODUCTION

Half of all vertebrate species are teleost fish (Nelson, 1994), the most speciose and diverse group of vertebrates (Fig. 1). What evolutionary mechanisms contributed to this remarkably successful explosion of biodiversity? A clue comes from the observation that chromosomally diploid teleosts often have several paralogous copies of single copy tetrapod genes (Morizot et al., 1991; Ekker et al., 1992, 1997; Akimenko et al., 1995; Smith et al., 2000, 2001; Adamska et al., 2001; Robinson-Rechavi et al., 2001). Genetic mapping studies showed that duplicated zebrafish genes map in duplicated chromosome segments co-orthologous to portions of individual human chromosomes (Amores et al., 1998; Postlethwait et al., 1998; Gates et al., 1999; Woods et al., 2000). The most parsimonious conclusion was that there had been a genome duplication in the zebrafish lineage, and subsequent analysis shows that this genome amplification in the ray-fin fish lineage occurred before the teleost radiation (Amores et al., 1998; Meyer and Schartl, 1999; Taylor et al., 2001a, b, 2003; Van de Peer et al., 2002; Amores et al., 2003). Did genome duplication play a role in the teleost radiation, and if so, how did it spur lineage diversification and morphologic variation?

To clarify the relationship between genomic and phenotypic complexity, we must first understand the processes by which duplicate genes evolve. Initial models of duplicate gene evolution assumed that each gene performs one or few functions (Ohno, 1970; Nei and Roychoudhury,

Institute of Neuroscience, University of Oregon, Eugene, Oregon
Grant sponsor: National Institutes of Health; Grant numbers: R01RR10715; 5 F32 GM020892; Grant sponsor: National Science Foundation; Grant numbers: IBN 0236239; IBN 9728587.
*Correspondence to: John H. Postlethwait, Institute of Neuroscience, University of Oregon, Eugene, Oregon, 97403-1254. E-mail: jpostle@uoneuro.uoregon.edu

DOI 10.1002/dvdy.10424

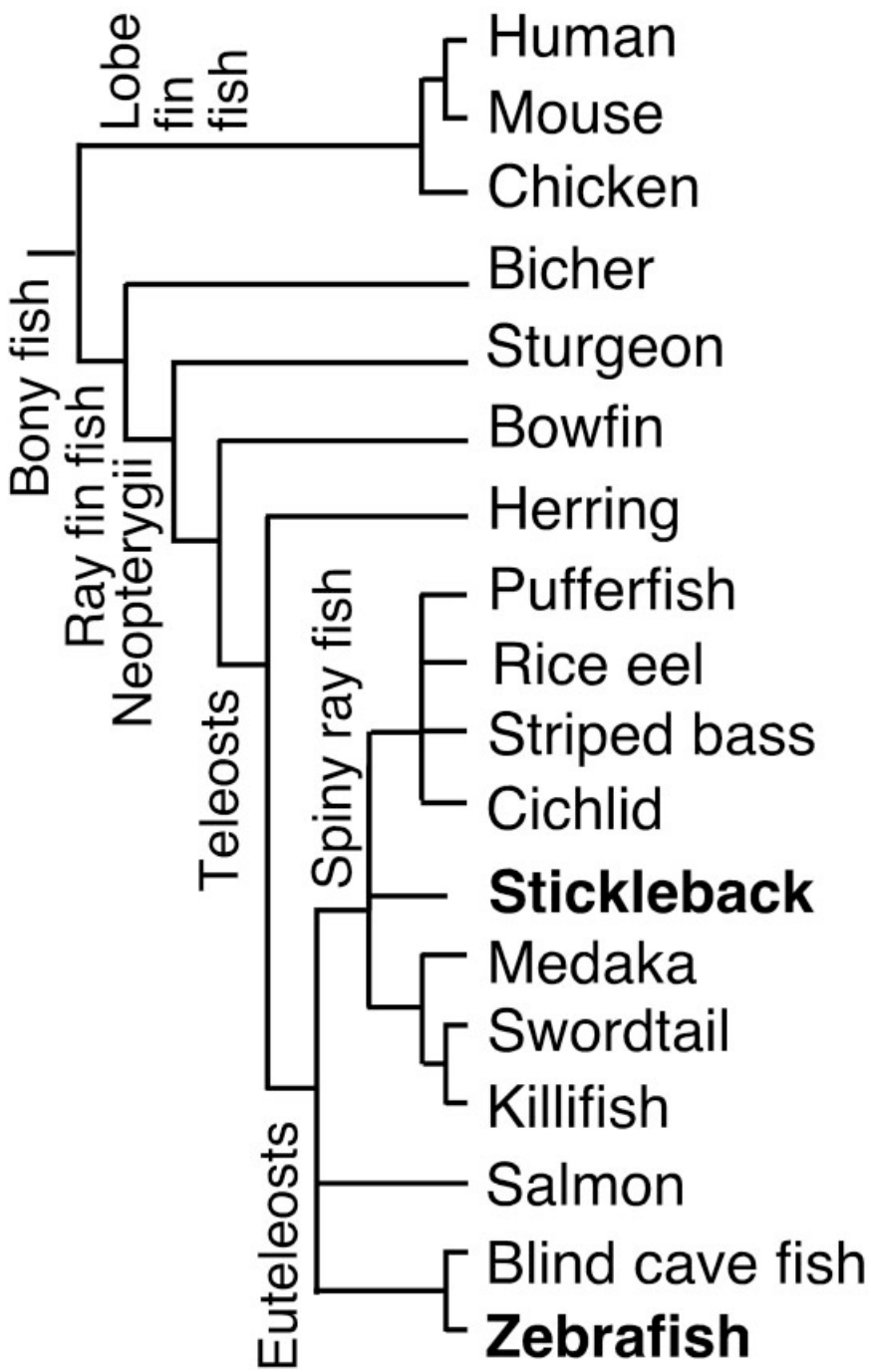

Fig. 1. A family tree for bony fish (Osteichthys) after (Nelson, 1994).

1973; Bailey et al., 1978; Takahata and Maruyama, 1979; Li, 1980; Watterson, 1983). On the basis of this assumption, retention of both copies of duplicated genes was hypothesized to be rare and would occur only if one copy acquired a novel, positively selected function (neofunctionalization; Walsh, 1995; Sidow, 1996; Cooke et al., 1997; Nadeau and Sankoff, 1997). Empirical data, however, show that duplicate pairs are retained at a higher rate than the classic models of duplicate gene evolution predict (Allendorf et al., 1975; Bisbee et al., 1977; Ferris and Whitt, 1979; Graf and Kobel, 1991; Hughes and Hughes, 1993). This may be due to the modular and complex nature of many genes, which comprise numerous subfunctions that perform a variety of tasks in different tissues and at different developmental times. This view of genes comprising multiple subfunctions has important implications for understanding both the evolution of duplicate gene pairs and the evolution of phenotypic complexity (Force et al., 2003). Instead of the acquisition of novel functions, the partitioning of ancestral subfunctions among descendant gene duplicates by the reciprocal, neutral fixation of degenerative regulatory mutations can contribute to permanent preservation of both copies (Hughes, 1994, 1999; Force et al., 1999; Hughes, 1999; Stoltzfus, 1999). The partitioning process for duplicate gene retention was formalized in the Duplication, Degeneration, Complementation (DDC) model (Force et al., 1999), and the partitioning of ancestral subfunctions has since been demonstrated in many duplicated teleost gene pairs (co-orthologs) in single lineages (De Martino et al., 2000; Bingham et al., 2001; Bruce et al., 2001; Chiang et al., 2001; Lister et al., 2001; McClintock et al., 2001, 2002; Altschmied et al., 2002; Yu et al., 2003).

A key question for understanding the evolution of gene duplicates and the role of gene duplication in the teleost radiation is whether ancestral subfunctions assorted before or after the divergence of various teleost lineages. Ancestral subfunctions that have not partitioned between duplicates before lineage divergence remain available for subsequent differential partitioning in different lineages, a process that can contribute to reproductive isolation (Lynch and Force, 2000). Subfunctions of paralogous duplicates in the different lineages will be largely identical if partitioning happens soon after duplication. If partitioning is slow, however, it may frequently occur independently in different lineages, and as a consequence, duplicate paralogs will retain different combinations of ancestral subfunctions in different teleost lineages. A central question is the following: how often does subfunction partitioning occur independently in different lineages? To date, there is no generalizable answer, because only a single case of duplicate gene pairs has been ana-

lyzed in detail in two different teleost lineages (Lister et al., 2001; Altschmied et al., 2002). In this case, subfunction partitioning appears to have occurred before the divergence of zebrafish, swordtail, and pufferfish lineages at the base of the teleost radiation (see Fig. 1).

To test the hypothesis that duplicate genes can be partitioned independently, we have examined expression patterns for two duplicates of *Sox9* in the threespine stickleback *Gasterosteus aculeatus* and the zebrafish *Danio rerio* lineages that diverged early in the teleost radiation (Fig. 1). Stickleback is an emerging model system for the study of the evolution of development, particularly for rapid morphologic changes in bony armor plates and defensive spines over short periods of time (Bell et al., 1993; Bell and Foster, 1994; Bell and Orti, 1994; Ahn and Gibson, 1999; Bell, 2001; Peichel et al., 2001; Gibson, 2002). Thus, these data are not only useful for macro-evolutionary studies across divergent lineages but also lay the groundwork for examining how microevolutionary change in cartilage and bone regulatory genes might contribute to the origin of phenotypic variation among stickleback populations.

Mutations in human, mouse, and zebrafish have shown that *Sox9* is an important gene for the regulation of cartilage formation and the subsequent development of cartilage-replacement bones (Wagner et al., 1994; Bi et al., 2001; Kist et al., 2002; Yan et al., 2002). Sox9 is a member of the Sox protein family of transcription factors, which contain a SRY-like HMG-box that binds and bends DNA (Ng et al., 1997; Wegner, 1999). In mammals, *Sox9* is involved in testis determination (Wagner et al., 1994; Vidal et al., 2001) and regulates cartilage formation by binding to a cis-regulatory sequence in *COL2A1*, the human type II collagen gene (Bell et al., 1997; Lefebvre et al., 1997). *Sox9* expression occurs in mesenchymal condensations before and during chondrogenesis, and the expression pattern mirrors that of *COL2A1* in tetrapods and teleosts (Zhao et al., 1997; Healy et al., 1999; Chiang et al., 2001; Yan et al., 2002). In addition to the critical role played in chondrogenesis, studies on *Xenopus* embryos indicate that reducing the production of Sox9 protein with morpholino antisense-oligonucleotides inhibits the formation of neural crest, the progenitors of craniofacial cartilage (Spokony et al., 2002). Thus, *Sox9* is a multifunctional gene, playing important roles in testis determination, the formation of neural crest, and chondrogenesis.

Zebrafish has two *Sox9* genes: *Sox9a* and *Sox9b* (Chiang et al., 2001). Gene phylogenies and genetic mapping show that these are co-orthologs of the tetrapod *Sox9* gene and likely arose in the ray-fin genome duplication event (Chiang et al., 2001). Importantly, expression patterns of these duplicates in zebrafish exhibit overlapping subsets of the tetrapod expression domains (Chiang et al., 2001; Yan et al., 2002) and together approximate the expression pattern of the *Sox9* gene in mouse, suggesting the partitioning of ancestral subfunctions. To determine whether subfunction partitioning occurred before or after the teleost radiation, we have cloned two *Sox9* genes from stickleback, established their phylogenetic relationships to tetrapod and zebrafish *Sox9* genes, and compared their expression patterns with those of zebrafish and tetrapods. We found that the event that produced duplicated teleost *Sox9* genes occurred before the divergence of stickleback and zebrafish lineages, and so did the partitioning of most subfunctions assayed by embryonic expression analysis. These results have important implications for the generalization of experimental results among teleost model systems, for the ways in which duplicated genes evolve, and for the mechanisms that generated the magnificent teleost radiation.

RESULTS AND DISCUSSION

Sox9 Duplication Occurred Before Divergence of Stickleback and Zebrafish Lineages

By using redundant primers designed to amplify the first exon of *Sox9*, we screened Fosmid genomic libraries for stickleback and found seven positive clones. Restriction enzyme analysis divided these clones into two distinct classes, and we chose two clones from each class to analyze in depth. A combination of directed and shotgun sequencing of the subclones produced two distinct genomic contigs 9908 and 6495 base pairs long that each contain a *Sox* gene with DNA sequence similarity to *Sox9*. Gene prediction analysis using GENSCAN (Burge and Karlin, 1998) showed that each gene comprises three exons and two introns and is predicted to encode translated products of 464 and 477 amino acids that have high sequence similarity to *Sox9*. The presumptive HMG domain at the N-terminus of the protein is well conserved between these predicted peptides and *Sox9* co-orthologs from teleosts (pufferfish, zebrafish, rice eel; Bagheri-Fam et al., 2001; Chiang et al., 2001; Zhou et al., 2002) and tetrapods (chick, mouse and human; Wright et al., 1993; Wagner et al., 1994; Healy et al., 1999). The C-termini of the predicted proteins, however, are much less conserved both between the stickleback paralogs and across the other orthologs.

A phylogenetic analysis of these sequences confirmed that stickleback have at least two different *Sox9* genes (Fig. 2). Both of these sequences fell squarely inside the *Sox9* clade with a very high bootstrap value (1,000 of 1,000), showing that they are not orthologs of the closely related *Sox8* and *Sox10* clades. Furthermore, one stickleback sequence clusters within the *Sox9a* clade with pufferfish, rice eel, and zebrafish (Chiang et al., 2001; Zhou et al., 2002), whereas the other falls within the *Sox9b* clade alongside pufferfish and rice eel (Bagheri-Fam et al., 2001; Zhou et al., 2002) with high bootstrap support (728 of 1000). Because zebrafish *sox9b* fell as an outgroup to the other teleost *Sox9* genes, it was important to be sure it was not a *Sox8*, *Sox9*, or *Sox10*; therefore, we identified a zebrafish *sox8* gene as EST fi23c10 (AW153579) isolated in the Washington University Zebrafish EST Project, and mapped it to LG3 in a region with conserved synteny with human chromosome

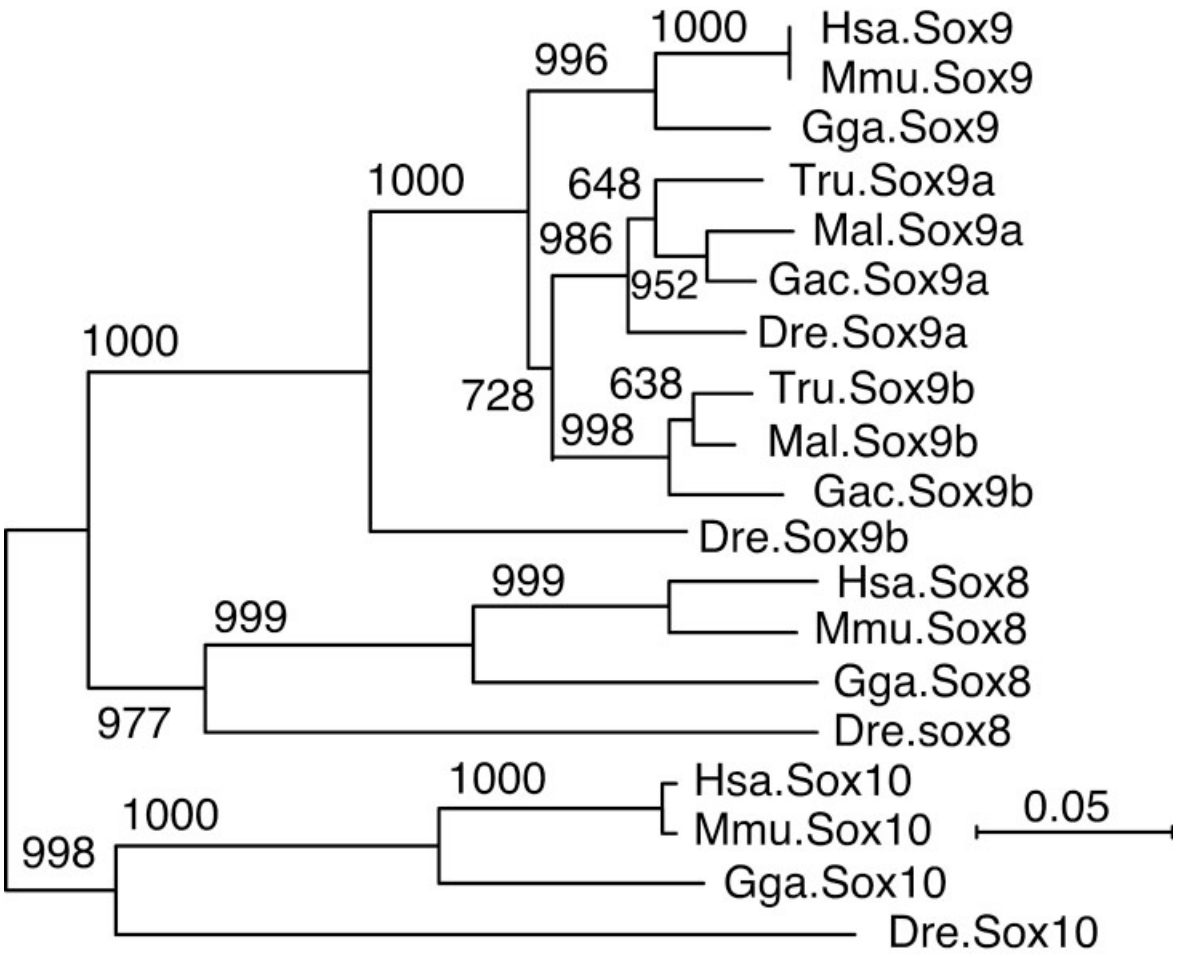

Fig. 2. A phylogenetic tree for *Sox9*-related genes. Numbers are bootstrap values for 1,000 trials. Dre, *Danio rerio*, zebrafish; Gac, *Gasterosteus aculeatus*, threespine stickleback; Gga, *Gallus gallus*, chicken; Hsa, *Homo sapiens*, human; Mal, *Monopterus albus*, rice eel; Mmu, *Mus musculus*, mouse; Tru, *Takifugu rubripes*, pufferfish. Dre Sox8 AW153579; Dre Sox9a AY090034; Dre Sox9b AAG09815; Dre Sox10 AF402677; Gga Sox9 U12533; Gga Sox10 AF152356; Gga Sox8 AF228664; Hsa Sox8 NP_055402; Hsa Sox9 NP_000337; Hsa Sox10 NP_008872; Mal Sox9a AF378150; Mal Sox9b AF378151; Mmu Sox8 XP_128601; Mmu Sox9 NP_035578; Mmu Sox10 XP_128139; Tru Sox9a AAL32172 (mayfold000587); Tru Sox9b mayfold 000421 (fugu assembly 3 at http://fugu.hgmp.mrc.ac.uk/).

Hsa16p13.3, the location of human *SOX8* (data not shown). This zebrafish sequence clusters strongly with the other vertebrate *Sox8* genes, which supports the conclusion that zebrafish *Sox9b* is a *Sox9* ortholog, despite its somewhat unexpected position in the tree. Importantly, the branching pattern of the *Sox9a* and *Sox9b* fish clades with respect to the tetrapod *Sox9* clade, while not completely unambiguous, shows that the duplication event that produced the teleost *Sox9a* and *Sox9b* clades occurred before the divergence of stickleback and zebrafish lineages. Because the zebrafish *Sox9a* and *Sox9b* genes map in duplicated chromosome segments that are co-orthologous to the human chromosome Hsa17 (Chiang et al., 2001), the location of *SOX9* (Wagner et al., 1994), we conclude that *Sox9a* and *Sox9b* arose in the hypothesized whole genome duplication event in the ray-fin fish lineage (Amores et al., 1998; Postlethwait et al., 1998; Taylor et al., 2001a, 2003). Thus, stickleback, zebrafish, pufferfish, and rice eel all appear to have retained both *Sox9* copies formed through the duplication of an ancestral fish genome approximately 300 million years ago.

Shared and Divergent Expression Patterns of *Sox9a*

Because stickleback and zebrafish *Sox9* genes were duplicated before the species lineages diverged, expression pattern analysis can distinguish between evolutionary changes that occurred after duplication but before lineage divergence and those that occurred between lineage divergence and the present. Stickleback embryos are cultured at 20°C and are larger and develop more slowly than zebrafish embryos cultured at 28.5°C, but the embryonic development of each is similar enough that appropriate comparisons can be made between embryos at morphologically similar developmental stages. Stickleback embryos at 32 hours postfertilization (h) show *Sox9a* expression in cells around the eye and otic placode (Fig. 3A). A zebrafish embryo at approximately the same stage of development shows strong *sox9a* expression in the region of the otic

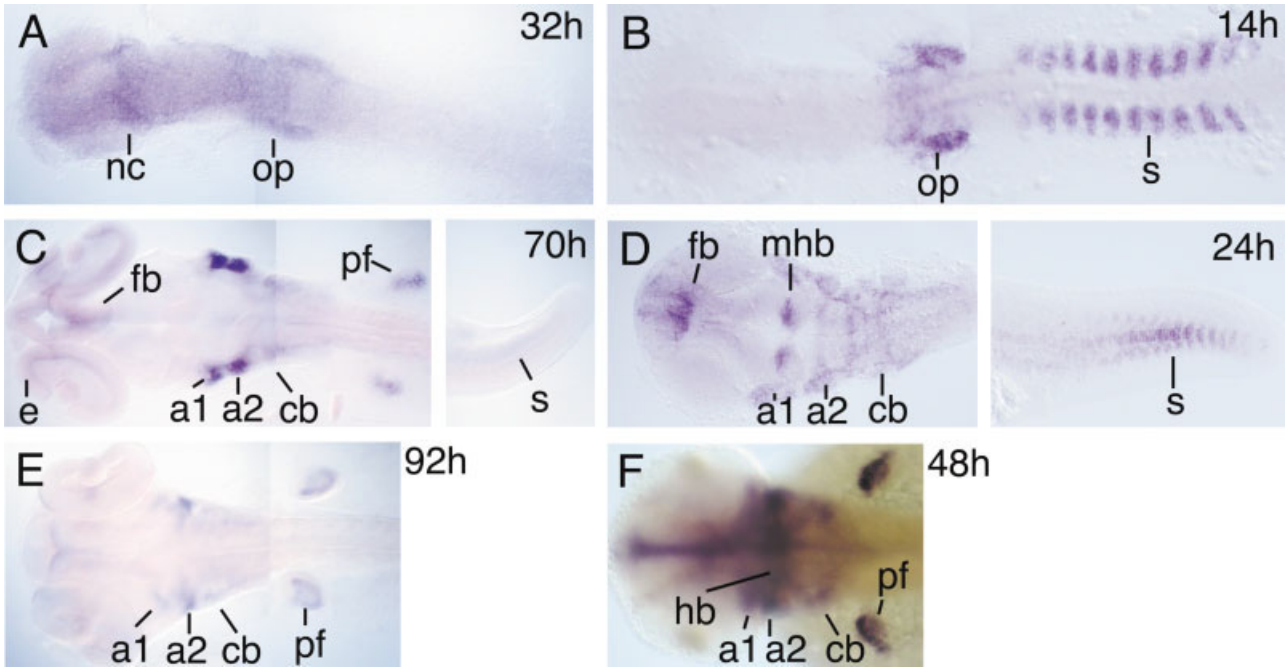

Fig. 3. *Sox9a* expression patterns. **A,C,E:** *Sox9a* expression in 32 hours postfertilization (h), 70-h, and 92-h stickleback embryos, respectively. **B,D,F:** *sox9a* expression in 14-h, 24-h, and 48-h zebrafish embryos. a1, mandibular arch; a2, hyoid arch; cb; ceratobranchial arches; e, eye; fb, forebrain; hb, hindbrain; mhb, midbrain–hindbrain border; nc, neural crest; op, otic placode region; pf, pectoral fin; s, somite.

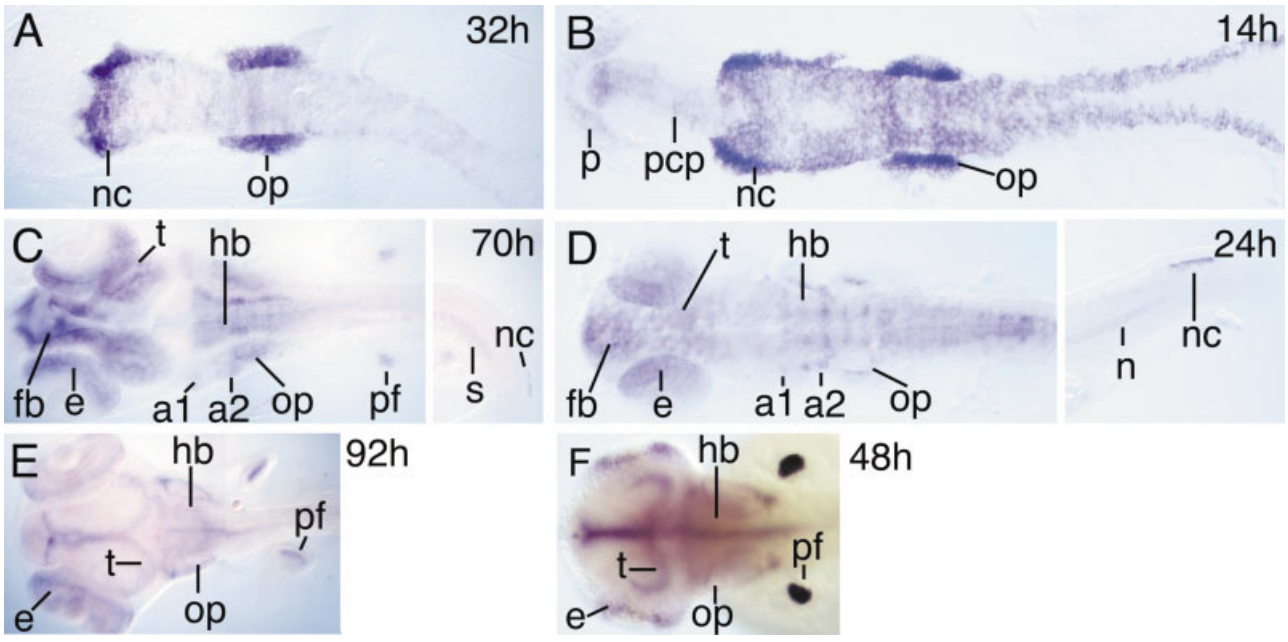

Fig. 4. *Sox9b* expression patterns. **A,C,E:** *Sox9a* expression in 32 hours postfertilization (h), 70-h, and 92-h stickleback embryos. **B,D,F:** *sox9a* expression in 14-h, 24-h, and 48-h zebrafish embryos. a1, mandibular arch; a2, hyoid arch; e, eye; fb, forebrain; hb, hindbrain; n. notochord; nc, neural crest; op, otic placode region; pcp, prechordal plate; pf, pectoral fin; s, somite; t, tectum.

placode and in the developing somites (Fig. 3B). Zebrafish *Sox9a* expression is weak around the eye, and reciprocally, stickleback *Sox9a* expression is weak in the somites, both expression domains appearing in this stage only after prolonged staining (data not shown).

By 70 h, stickleback embryos continue to have staining in the eye region, as well as in the forebrain (Fig. 3C). At this time, crest cells populating the pharyngeal arches, perhaps derived from the *Sox9a*-positive cells in the otic region, stain intensely, and expression has begun in the pectoral fin bud. Expression is particularly strong in the first (mandibular)

and second (hyoid) arches and weaker in the ceratobranchial arches. Weak expression of *Sox9a* appears in the tail somites in 70-h embryos (Fig. 3C). Most of these expression domains are similar in zebrafish embryos of approximately the same stage of development, with strong expression in forebrain, crest cells populating the arches, tail somites (Fig. 3D), and pectoral fin (data not shown). Two major differences occur in the midbrain–hindbrain border and the somites, where zebrafish has significantly stronger expression than stickleback.

In 92-h stickleback embryos, *Sox9a* continues to be expressed in the pharyngeal arches, and in the outer mesenchyme of the pectoral fin, more strongly in the outer portion of the mesenchyme than the central core, as it is in zebrafish (Fig. 3E,F). In zebrafish, *Sox9a* is expressed in a striped pattern in the hindbrain, likely in glial cells as it is in mouse (Pompolo and Harley, 2001). These results show that the patterns of *Sox9a* expression in stickleback and zebrafish are largely the same, but differ in detail, with the stickleback gene stronger in the eye region, and the zebrafish gene stronger in the somites and midbrain–hindbrain border. Expression domains that are taxon-specific involve evolutionary change in developmental regulation after the divergence of stickleback and zebrafish lineages.

Shared and Divergent Expression Patterns of *Sox9b*

Like *Sox9a*, *Sox9b* is expressed in generally similar patterns in stickleback and zebrafish. At 32 h, stickleback embryos express *Sox9b* around the eye and otic vesicle, and in the neural crest in the head and trunk (Fig. 4A). This stickleback expression pattern is largely the same as the pattern for the orthologous gene in zebrafish (Fig. 4B), with the main difference being that crest expression in the trunk is comparatively stronger in zebrafish.

In 70-h stickleback embryos, the expression pattern of *Sox9b* is more complex and extensive than *Sox9a* in the cranial region. At this stage, *Sox9b* transcripts accumulate in the retina of the eye and in the forebrain and tectum (Fig. 4C). In the hindbrain, *Sox9b* is expressed in both teleosts in a striped pattern as for *Sox9a* in zebrafish (Fig. 3F). *Sox9b* is expressed in the pharyngeal arches more strongly in the ceratobranchials than in the mandibular and hyoid arches in both teleost embryos. At this stage, *Sox9b* is expressed weakly in the somites of stickleback but not zebrafish embryos, and, like zebrafish, *Sox9b* is expressed in the crest at the end of the tail (Fig. 4C,D).

By 92 h, expression of *Sox9b* in the stickleback brain is mainly in the peripheral parts of the tectum and in cells lining the ventricles (Fig. 4E), as in zebrafish (Fig. 4F). In the fin bud, *Sox9b* transcripts accumulate in the central core of the mesenchyme in both stickleback and zebrafish. In general, the expression patterns of *Sox9b* are more similar between the two teleosts than are the expression patterns of *Sox9a*.

Conclusions and Future Directions

Stickleback and zebrafish embryos have largely similar embryonic expression patterns for *Sox9a* and *Sox9b*, and these combined *Sox9a* and *Sox9b* domains are generally possessed by tetrapod outgroups as a function of a single *Sox9* ortholog, suggesting that they are ancestral. Most partitioned expression differences between the duplicate genes are common to both teleost species, such as strong expression in the mandibular and hyoid arches for *Sox9a* and strong expression in the trunk crest for *Sox9b*. These cases display the behavior expected if ancestral regulatory elements driving these expression domains reciprocally partitioned between the duplicate genes in the time interval between the gene duplication event and the divergence of stickleback and zebrafish lineages. This is also the case for *Mitf* gene duplicates in teleosts. In tetrapods, there is a single *Mitf* gene with transcripts that are expressed from different promotors and 5′ exons (Yasumoto et al., 1998; Udono et al., 2000). In teleosts, in contrast, there are duplicated genes *mitfa* and *mitfb*, which produce these alternative isoforms expressed in general in similar patterns to their tetrapod orthologs (Lister et al., 2001; Altschmied et al., 2002). Thus, for the only two cases studied, the general result is that most ancestral gene subfunctions appear to have partitioned before the teleost radiation. Clearly more cases need to be examined in detail before a strong generalization can be made.

For some ancestral expression domains, however, each co-ortholog is expressed in a species-specific manner. Examples include the strong expression of *Sox9a* in the zebrafish midbrain–hindbrain border and somites, but undetected or weak expression of *Sox9a* in these domains in stickleback, and reciprocally, the strong expression of *Sox9b* around the eye in stickleback but the weak or absent expression of *sox9b* in the same region of zebrafish embryos. In general, for the *Sox9* gene pair, the species-specific features appear to be mainly quantitative rather than qualitative. Such unshared domains have the features predicted for regulatory subfunctions that had not partitioned at the time of the stickleback/zebrafish lineage divergence but then evolved differently in the two lineages.

A third class of expression patterns appears to be shared by *Sox9a* and *Sox9b* in both teleost embryos. Expression in the hindbrain and portions of the pectoral fin mesenchyme appear to be examples. If higher resolution cell-by-cell analysis shows that individual cells are transcribing both duplicates at the same time, then these would be examples of ancestral regulatory elements that remain unpartitioned to the present. However, the possibility remains that, even though at a gross level of examinations the domains of expression appear to be overlapping, a more detailed analysis might reveal fine scale partitioning across cells within structures such as the fin mesenchyme. In-depth studies of apparently overlapping domains should be completed before conclusions are drawn about whether domains remain unpartitioned in these tissues.

The results with *Sox9* and *Mitf*

genes suggest that most gene subfunctions may have assorted between duplicated genes before the teleost explosion. But did the partitioning happen rapidly after duplication or did it take a long period of time relative to the duplication event and the present? An answer depends upon knowing the date of the genome duplication and the dates at which teleost lineages diverged. Unfortunately, the dating of both events is in question. Molecular clock analysis has suggested that the ray-fin fish genome duplication occurred more than 300 million years ago (MYA) (Taylor et al., 2001a) and that the teleost radiation occurred approximately 140 MYA (Hedges and Kumar, 2002), but paleontologic evidence suggests that the radiation may have occurred 200 MYA or more (Santini and Tyler, 1999). Genomic analysis is needed for basally diverging teleosts such as eels, and basally diverging ray-fin fish, such as bowfin and bicher, to further resolve the question. At the extreme, however, the times between genome duplication and the teleost radiation, and the radiation to the present, were about equal—approximately 150 million years. However, we hypothesize that the teleost radiation may have occurred sooner after the genome duplication than previously thought, with perhaps only 50–100 million years intervening between the two. In either timing scenario, however, most subfunctions appear to have partitioned before the teleost radiation, according to currently available results. This timing would support the idea that subfunction partitioning after duplication may be a relatively rapid process.

The partitioning of ancestral subfunctions among duplicated genes may contribute to phenotypic evolution. Darwin said that it is generally acknowledged that all organic beings have been formed on two great laws - Unity of Type and Conditions of Existence (Darwin, 1859). The results of subfunction partitioning in teleost *Sox9* co-orthologs suggests the hypothesis that subfunction partitioning may contribute mechanistically to Darwin's generalization. The early subfunction partitioning that is shared across lineages could provide the unity of type, whereas later lineage-specific, largely quantitative partitioning, could provide the genetic fodder that allows lineages to acquire different distributions of traits appropriate for their conditions of existence, and thus their divergence. Because genome duplication provides a genome's worth of gene duplicates, even within a largely shared set of partitioning that happened soon after duplication, it would still offer enormous opportunities for subsequent independent subfunction partitioning and multiple dimensions along which lineages could diverge.

The above considerations are based on a very small data set. What is now needed is developmental genetic analysis of a large number of gene duplicates in a variety of different teleosts, including models for microevolution, such as stickleback and swordtails (Walter and Kazianis, 2001) and models amenable to mutagenesis such as zebrafish and medaka (Grunwald and Streisinger, 1992; Loosli et al., 2000) to understand the relative frequency of conserved vs. independent partitioning, and what role independent partitioning plays in divergence. On a practical note, establishing this frequency is very important because it will provide a sense of how often we should expect model systems that have experienced a genome duplication event, such as pufferfish, medaka, and zebrafish, to unambiguously provide insight into the function of genes in other organisms, particularly humans. At a more fundamental level, the hypothesis that subfunction partitioning played a role in the teleost radiation makes the testable prediction that lineages with the most independent partitioning should be the most phenotypically diverse. Ruling out this hypothesis would focus attention on other causes for the most spectacular vertebrate example of biodiversity.

EXPERIMENTAL PROCEDURES

Isolation of Stickleback *Sox9a* and *Sox9b*

We constructed two stickleback Fosmid genomic libraries by using the CopyControl Fosmid Library Production Kit from Epicentre (catalog no. CCFOS110). Each library was made from a single individual, both collected from the wild in Alaska. One individual, collected from Rabbit Slough, was from an anadromous population exhibiting extensive bony lateral plate and pelvic armor. The other individual, collected from Bear Paw Lake, was from a population of stickleback whose members have lost almost all bony armor. These fish were transported to the University of Oregon stickleback facility and reared for 2 months before being killed. Stickleback clone inserts average approximately 50 kbp in length, and the arrayed libraries provide approximately 12× coverage of the stickleback genome, assuming 600 Mbp in the haploid genome (Vinogradov, 1998).

We screened the library for *Sox9* genes by using polymerase chain reaction (PCR) primers designed to conserved regions of the first exons of the *Fugu rubripes Sox9a* and *Sox9b* genes. The forward primer was 5′-TGAATCTCCTCGACCCTTACC-3′, and the reverse primer was 5′-TGCAGCCTGAGCCCACAC-3′. Seven independent positive clones were identified. We sheared and subcloned four fosmid clones that were both positive for *Sox9*, but which had two distinct restriction enzyme digestion patterns. Shearing was performed by using a nebulizer and the ends of the subclones were repaired by using a mixture of T4 polymerase and Klenow fragment. The blunted subclones were then ligated into the PCR4-BLUNT cloning vector (Invitrogen TOPO Shotgun Subcloning Kit, catalog nos. K7000-01, K7010-01, K7050-01, and K7060-01). The average size of inserted DNA was 2–5 kb. Subclones were arrayed into 96-well plates as single clones and then screened by means of PCR with the same first exon primers described above. Positive clones were sequenced from both ends of the vector. Additionally, 16 randomly chosen subclones from each Fosmid were sequenced. In all, this method provided full-length genomic sequence for Sox9a (9908 bp) and *Sox9b* (6495 bp), including 5′ and 3′ untranslated regions (UTRs), introns,

and perhaps some regulatory elements. Each *Sox9*-positive contig was sequenced to an average of 6× coverage, equally distributed in either direction (GenBank accession nos. AY351914 and AY351915). The gene structure of the *Sox9* gene in each contig was predicted by using the GENSCAN (Burge and Karlin, 1998) Web server (http://genes.mit.edu/GENSCAN.html).

Phylogenetic Analysis

Protein sequence for zebrafish, rice eel, mouse, human, and chicken orthologs of *Sox8*, *Sox9*, and *Sox10* were obtained from Entrez (http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?db=Nucleotide). The zebrafish *Sox9a* and *Sox9b* sequences were used to BLAST (Altschul et al., 1997) against the Fugu database (http://ensembl.fugu-sg.org/Fugu_rubripes/blastview), and we obtained one single strong alignment for each paralog. The gene structure for each was determined by analyzing the genomic sequence using the GENSCAN Web server, as was done for the stickleback genes. The published *Sox9* protein sequences and the inferred pufferfish and stickleback protein sequences were analyzed by using ClustalX (version 1.82; Thompson et al., 1994) to establish the phylogenetic relationship of these sequences. Node confidence was ascertained by bootstrapping the data set 1,000 times (Felsenstein, 1985), and the resulting tree was visualized by using NJPlot (Perrière and Gouy, 1996; http://pbil.univ-lyon1.fr/software/njplot.html).

Detecting Gene Expression in Stickleback Embryos by In Situ Hybridization

Linearized genomic clones of each gene were transcribed in vitro to make riboprobes using digoxigenin-labeled UTP. The *Sox9a* probe was generated from a subclone in a PCR4-Shotgun/Not1-linearized template using T3 RNA polymerase, whereas the *Sox9b* probe was synthesized from a PCR4-TOPO/Not1-linearized plasmid also using T3 RNA polymerase. The Sox9a probe is 650 bp in length, covering the 5′UTR and most of the first exon. The *Sox9b* probe is 1,800 bp in length, covering most of exons 1 and 3, and all of exon 1 and introns 1 and 2. Embryos were fixed in 4% paraformaldehyde at 20°C for 1 week, after which time they were dechorionated by hand and processed for in situ hybridization as described (Chiang et al., 2001; Yan et al., 2002). Stickleback embryos were staged based on morphologic criteria in analogy with the zebrafish staging series (Kimmel et al., 1995).

ACKNOWLEDGMENTS

We thank for support NSF grants IBN 0236239 for stickleback, NSF grant IBN 9728587 for sex gene research, NIH grant R01RR10715 for zebrafish, and NIH grant 5 F32 GM020892 for postdoctoral training support (W.A.C.). NSF IGERT grant DGE 9972830 supported summer undergraduate researchers Diana Bradley, Rebecca Loda, Melanie Robinson, and Mark Rothgary, who helped screen libraries. We also thank Mike Bell, Susan Foster, John Baker, and Jeff Walker and their students for developing the system of Alaskan stickleback populations into the great resource for studies in microevolution and development that it is today.

REFERENCES

Adamska M, Wolff A, Kreusler M, Wittbrodt J, Braun T, Bober E. 2001. Five Nkx5 genes show differential expression patterns in anlagen of sensory organs in medaka: insight into the evolution of the gene family. Dev Genes Evol 211:338–349.

Ahn D, Gibson G. 1999. Expression patterns of threespine stickleback Hox genes and insights into the evolution of the vertebrate body axis. Dev Genes Evol 209:482–494.

Akimenko M-A, Johnson SL, Westerfield M, Ekker M. 1995. Differential induction of four *msx* homeobox genes during fin development and regeneration in zebrafish. Development 121:347–357.

Allendorf FW, Utter FM, May BP. 1975. Gene duplication within the family Salmonidae: II. Detection and determination of the genetic control of duplicate loci through inheritance studies and the examination of populations. In: Markert CL, editor. Isozymes. New York: Academic Press. p 415–432.

Altschmied J, Delfgaauw J, Wilde B, Duschl J, Bouneau L, Volff JN, Schartl M. 2002. Subfunctionalization of duplicate mitf genes associated with differential degeneration of alternative exons in fish. Genetics 161:259–267.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402.

Amores A, Force A, Yan Y-L, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang Y-L, 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 J. 2003. Developmental roles of pufferfish Hox clusters and genome evolution in ray-fin fish. Genome Research (in press).

Bagheri-Fam S, Ferraz C, Daemaille J, Scherer G, Pfeifer D. 2001. Comparative genomics of the SOX9 region in human and *Fugu rubripes*: conservation of short regulatory sequence elements within large intergenic regions. Genomics 78:73–82.

Bailey GS, Poulter RTM, Stockwell PA. 1978. Gene duplication in tetraploid fish: model for gene silencing at unlinked duplicated loci. Proc Natl Acad Sci U S A 75:5575–5579.

Bell MA. 2001. Ontogeny and evolution of lateral plate number in low morph threespine stickleback fish, Gasterosteus aculeatus. J Morphol 248:205.

Bell MA, Foster SA. 1994. The evolutionary biology of the threespine stickleback. Oxford: Oxford University Press. 571 p.

Bell MA, Orti G. 1994. Pelvic reduction in threespine stickleback from Cook Inlet lakes: geographical distribution and intrapopulation variation. Copeia:314–325.

Bell MA, Orti G, Walker JA, Koenings JP. 1993. Evolution of pelvic reduction in threespine stickleback fish: a test of competing hypotheses. Evolution 47: 906–914.

Bell DM, Leung KK, Wheatley SC, Ng L-J, Zhou S, Ling KW, Sham MH, Koopman P, Tam PPL, Cheah KSE. 1997. SOX9 directly regulates the type-II collagen gene. Nat Genet 16:174–178.

Bi W, Huang W, Whitworth DJ, Deng JM, Zhang Z, Behringer RR, de Crombrugghe B. 2001. Haploinsufficiency of *Sox9* results in defective cartilage primordia and premature skeletal mineralization. Proc Natl Acad Sci U S A 98:6698–6703.

Bingham S, Nasevicius A, Ekker SC, Chandrasekhar A. 2001. Sonic hedgehog and tiggy-winkle hedgehog cooperatively induce zebrafish branchiomotor neurons. Genesis 30:170–174.

Bisbee CA, Baker MA, Wilson AC, Irandokht HA, Fischberg M. 1977. Albumin phylogeny for clawed frogs (*Xenopus*). Science 195:785–787.

Bruce AE, Oates AC, Prince VE, Ho RK. 2001. Additional hox clusters in the zebrafish: divergent expression patterns

belie equivalent activities of duplicate *hoxB5* genes. Evol Dev 3:127–144.

Burge CB, Karlin S. 1998. Finding the genes in genomic DNA. Curr Opin Struct Biol 8:346–354.

Chiang E, Pai C-I, Wyatt M, Yan Y-L, Postlethwait J, Chung B-C. 2001. Two *sox9* genes on duplicated zebrafish chromosomes: expression of similar transcription activators in distinct sites. Dev Biol 229:149–163.

Cooke JMA, Nowak MA, Boerlijst M, Maynard Smith J. 1997. Evolutionary origins and maintenance of redundant gene expression during metazoan development. Trends Genet 13:360–364.

Darwin C. 1859. The origin of species. London: John Murray.

De Martino S, Yan Y-L, Jowett T, Postlethwait JH, Varga ZM, Ashworth A, Austin CA. 2000. Expression of *sox11* gene duplicates in zebrafish suggests the reciprocal loss of ancestral gene expression patterns in development. Dev Dyn 217: 279–292.

Ekker M, Wegner J, Akimenko M-A, Westerfield M. 1992. Coordinate embryonic expression of three zebrafish *engrailed* genes. Development 116: 1001–1010.

Ekker M, Akimenko M, Allende M, Smith R, Drouin G, Langille R, Weinberg E, Westerfield M. 1997. Relationships among *msx* gene structure and function in zebrafish and other vertebrates. Mol Biol Evol 14:1008–1022.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791.

Ferris SD, Whitt GS. 1979. Evolution of the differential regulation of duplicate genes after polyploidization. J Mol Evol 12:267–317.

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

Force AG, Cresko WA, Pickett FB. 2003. Informational accretion, gene duplication, and the mechanisms of genetic module parcellation. In: Schlosser G, Wagner G, editors. Modularity in development and evolution. Chicago: University of Chicago Press.

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.

Gibson G. 2002. A genetic attack on the defense complex. Bioessays 24:487–489.

Graf J-D, Kobel HR. 1991. *Xenopus laevis*: practical uses in cell and molecular biology: genetics. In: Kay BK, Peng HB, editors. Methods in cell biology. New York: Academic Press. p 19–34.

Grunwald DJ, Streisinger G. 1992. Induction of recessive lethal and specific locus mutations in the zebrafish with ethyl nitrosourea. Genet Res 59:103–116.

Healy C, Uwanogho D, Sharpe PT. 1999. Regulation and role of Sox9 in cartilage formation. Dev Dyn 215:69–78.

Hedges SB, Kumar S. 2002. Genomics. Vertebrate genomes compared. Science 297:1283–1285.

Hughes AL. 1994. The evolution of functionally novel proteins after gene duplication. Proc R Soc Lond B 256:119–124.

Hughes AL. 1999. Adaptive evolution of genes and genomes. New York: Oxford University Press.

Hughes MK, Hughes AL. 1993. Evolution of duplicate genes in a tetraploid animal, *Xenopus laevis*. Mol Biol Evol 10:1360–1369.

Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF. 1995. Stages of embryonic development of the zebrafish. Dev Dyn 203:253–310.

Kist R, Schrewe H, Balling R, Scherer G. 2002. Conditional inactivation of *Sox9*: a mouse model for campomelic dysplasia. Genesis 32:121–123.

Lefebvre V, Huang W, Harley VR, Goodfellow PN, de Crombrugghe B. 1997. SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro alpha1(II) collagen gene. Mol Cell Biol 17:2336–2346.

Li W-H. 1980. Rate of gene silencing at duplicate loci: a theoretical study and interpretation of data from tetraploid fishes. Genetics 95:237–258.

Lister J, Close J, Raible D. 2001. Duplicate *mitf* genes in zebrafish: complementary expression and conservation of melanogenic potential. Dev Biol 237: 333–344.

Loosli F, Koster RW, Carl M, Kuhnlein R, Henrich T, Mucke M, Krone A, Wittbrodt J. 2000. A genetic screen for mutations affecting embryonic development in medaka fish (Oryzias latipes). Mech Dev 97:133–139.

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

McClintock JM, Carlson R, Mann DM, Prince VE. 2001. Consequences of *Hox* gene duplication in the vertebrates: an investigation of the zebrafish *Hox* paralogue group 1 genes. Development 128: 2471–2484.

McClintock JM, Kheirbek MA, Prince VE. 2002. Knockdown of duplicated zebrafish *hoxb1* genes reveals distinct roles in hindbrain patterning and a novel mechanism of duplicate gene retention. Development 129:2339–2354.

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.

Morizot DC, Slaugenhaupt SA, Kallman KD, Chakravarti A. 1991. Genetic linkage map of fishes of the genus *Xiphophorus* (Teleostei: Poeciliidae). Genetics 127:399–410.

Nadeau JH, Sankoff D. 1997. Comparable rates of gene loss and functional divergence after genome duplications early in vertebrate evolution. Genetics 147:1259–1266.

Nei M, Roychoudhury AK. 1973. Probability of fixation of nonfunctional genes at duplicate loci. Am Nat 107:362–372.

Nelson JS. 1994. Fishes of the world. New York: Wiley-Interscience.

Ng L-J, Wheatley S, Muscat GEO, Conway-Campbell J, Bowles J, Wright E, Bell DM, Tam PPL, Cheah KSE, Koopman P. 1997. SOX9 binds DNA, activates transcription, and coexpresses with type II collagen during chondrogenesis in the mouse . Dev Biol 183:108–121.

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

Peichel CL, Nereng KS, Ohgi KA, Cole BL, Colosimo PF, Buerkle CA, Schluter D, Kingsley DM. 2001. The genetic architecture of divergence between threespine stickleback species. Nature 414:901–905.

Perrière G, Gouy M. 1996. WWW-query: an on-line retrieval system for biological sequence banks. Biochimie 78:364–369.

Pompolo S, Harley VR. 2001. Localisation of the SRY-related HMG box protein, SOX9, in rodent brain. Brain Res 906:143–148.

Postlethwait J, Yan Y, Gates M, Horne S, Amores A, Brownlie A, Donovan A, Egan E, Force A, Gong Z, Goutel C, Fritz A, Kelsh R, Knapik E, Liao E, Paw B, Ransom D, Singer A, Thomson M, Abduljabbar T, Yelick P, Beier D, Joly J, Larhammar D, Talbot W. 1998. Vertebrate genome evolution and the zebrafish gene map. Nat Genet 18:345–349.

Robinson-Rechavi M, Marchand O, Escriva H, Bardet PL, Zelus D, Hughes S, Laudet V. 2001. Euteleost fish genomes are characterized by expansion of gene families. Genome Res 11:781–788.

Santini F, Tyler JC. 1999. A new phylogenetic hypothesis for the order Tetraodontiformes (Teleostei, Pisces), with placement of the most fossil basal lineages. Am Zool 39:10A.

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

Smith S, Metcalfe JA, Elgar G. 2000. Identification and analysis of two snail genes in the pufferfish (Fugu rubripes) and mapping of human SNA to 20q. Gene 247:119–128.

Smith S, Metcalfe JA, Elgar G. 2001. Characterization of two topoisomerase 1 genes in the pufferfish (*Fugu rubripes*). Gene 265:195–204.

Spokony R, Aoki Y, Saint-Germain N, Magner-Fink E, Saint-Jeannet J. 2002. The transcription factor *Sox9* is required for cranial neural crest development in *Xenopus*. Development 129:421–432.

Stoltzfus A. 1999. On the possibility of constructive neutral evolution. J Mol Evol 49:169–181.

Takahata N, Maruyama T. 1979. Polymorphism and loss of duplicate gene ex-

pression: a theoretical study with application to tetraploid fish. Proc Natl Acad Sci U S A 76:4521–4525.

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

Taylor JS, Van de Peer Y, Meyer A. 2001b. Revisiting recent challenges to the ancient fish-specific genome duplication hypothesis. Curr Biol 11:R1005–R1008.

Taylor J, 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, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive and multiple sequence alignment through sequence weighting, positions-specific gap penalties, and weight matrix choice. Nucleic Acids Res 22:4673–4680.

Udono T, Yasumoto K, Takeda K, Amae S, Watanabe K, Saito H, Fuse N, Tachibana M, Takahashi K, Tamai M, Shibahara S. 2000. Structural organization of the human microphthalmia-associated transcription factor gene containing four alternative promoters. Biochim Biophys Acta 1491:205–219.

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

Vidal VPI, Chaboissier MC, de Rooij DG, Schedl A. 2001. *Sox9* induces testis development in XX transgenic mice. Nat Genet 28:216–217.

Vinogradov A. 1998. Genome size and GC-percent in vertebrates as determined by flow cytometry: the triangular relationship. Cytometry 31:100–109.

Wagner T, Wirth J, Meyer J, Zabel B, Held M, Zimmer J, Pasantes J, Bricarelli FD, Keutel J, Hustert E. 1994. Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the *SRY*-related gene *SOX9*. Cell 79:111–1120.

Walsh JB. 1995. How often do duplicated genes evolve new functions? Genetics 110:345–364.

Walter RB, Kazianis S. 2001. Xiphophorus interspecies hybrids as genetic models of induced neoplasia. ILAR J 42:299–321.

Watterson GA. 1983. On the time for gene silencing at duplicate loci. Genetics 105:745–766.

Wegner M. 1999. From head to toes: the multiple facets of Sox proteins. Nucleic Acids Res 27:1409–1420.

Woods IG, Kelly PD, Chu F, Ngo-Hazelett P, Yan Y-L, Huang H, Postlethwait JH, Talbot WS. 2000. A comparative map of the zebrafish genome. Genome Res 10:1903–1914.

Wright EM, Snopek B, Koopman P. 1993. Seven new members of the *Sox* gene family expressed during mouse development. Nucleic Acids Res 21:744.

Yan YL, Miller CT, Nissen R, Singer A, Liu D, Kirn A, Draper B, Willoughby J, Morcos PA, Amsterdam A, Chung BC, Westerfield M, Haffter P, Hopkins N, Kimmel C, Postlethwait JH. 2002. A zebrafish sox9 gene required for cartilage morphogenesis. Development 129:5065–5079.

Yasumoto K, Amae S, Udono T, Fuse N, Takeda K, Shibahara S. 1998. A big gene linked to small eyes encodes multiple Mitf isoforms: many promoters make light work. Pigment Cell Res 11: 329–336.

Yu WP, Brenner S, Venkatesh B. 2003. Duplication, degeneration and subfunctionalization of the nested synapsin-Timp genes in Fugu. Trends Genet 19: 180–183.

Zhao Q, Eberspaecher H, Lefebvre V, de Crombrugghe B. 1997. Parallel expression of *Sox9* and *Col2a1* in cells undergoing chondrogenesis. Dev Dyn 209: 377–386.

Zhou R, Cheng H, Zhang Q, Guo Y, Cooper R, Tiersch T. 2002. SRY-related genes in the genome of the rice field eel (Monopterus albus). Genet Sel Evol 34:129–137.