

Hox Cluster Duplication in the Basal Teleost *Hiodon alosoides* (Osteoglossomorpha)

Karen E. Chambers
Ryan McDaniel
Jeremy D. Raincrow
Maya Deshmukh
Peter F. Stadler

SFI WORKING PAPER: 2008-09-041

SFI Working Papers contain accounts of scientific work of the author(s) and do not necessarily represent the views of the Santa Fe Institute. We accept papers intended for publication in peer-reviewed journals or proceedings volumes, but not papers that have already appeared in print. Except for papers by our external faculty, papers must be based on work done at SFI, inspired by an invited visit to or collaboration at SFI, or funded by an SFI grant.

©NOTICE: This working paper is included by permission of the contributing author(s) as a means to ensure timely distribution of the scholarly and technical work on a non-commercial basis. Copyright and all rights therein are maintained by the author(s). It is understood that all persons copying this information will adhere to the terms and constraints invoked by each author's copyright. These works may be reposted only with the explicit permission of the copyright holder.

www.santafe.edu



SANTA FE INSTITUTE

Hox cluster duplication in the basal teleost *Hiodon alosoides* (Osteoglossomorpha)

Karen E. Chambers · Ryan McDaniel · Jeremy D. Raincrow · Maya Deshmukh · Peter F Stadler · Chi-hua Chiu

Manuscript date Thu Aug 14 02:02:53 CEST 2008

Abstract Large-scale – even genome-wide – duplications have repeatedly been invoked as an explanation for major radiations. Teleosts, the most species-rich vertebrate clade, underwent a “fish-specific genome duplication” (FSGD) that is shared by most ray-finned fish lineages. We investigate here the *Hox* complement of the goldeye (*Hiodon alosoides*), a representative of Osteoglossomorpha, the most basal teleostean clade. An extensive PCR survey reveals that goldeye has at least eight *Hox* clusters, indicating a duplicated genome compared to basal actinopterygians. The possession of duplicated *Hox* clusters is uncoupled to species richness. The *Hox* system of the goldeye is substantially different from that of other teleost lineages, having retained several duplicates of *Hox* genes for which crown teleosts have lost at least one copy. A detailed analysis of the PCR fragments as well as full length sequences of two *HoxA13* paralogs, and *HoxA10* and *HoxC4* genes places the duplication event close in time to the divergence of Osteoglossomorpha and crown teleosts. The data are consistent with — but do not conclusively prove — that Osteoglossomorpha shares the FSGD.

Keywords *Hox* clusters, Fish-Specific Genome Duplication, goldeye *Hiodon alosoides*

1 Introduction

Genome duplication is a powerful evolutionary mechanism that has contributed to the diversity of the vertebrate lineage (Ohno, 1970). Present evidence supports that two rounds of genome duplication (called 1R and 2R) occurred in early chordate phylogeny and are common to the ancestor of jawed vertebrates (cartilaginous, lobe-finned, and ray-finned fishes) (Sidow, 1996). The clade of ray-finned fishes (Actinopterygii, Figure 1) underwent a third round of genome duplication dubbed the 3R or the FSGD (fish specific genome duplication, red arrow in Figure 1) (Taylor *et al.*, 2001; Christoffels *et al.*, 2004; Vandepoele *et al.*, 2004). The FSGD is proposed to be a whole genome event (Taylor *et al.*, 2003; Brunet *et al.*, 2006), a fact that is well supported by the observation that spotted green pufferfish (Teleostei; *Tetraodon nigroviridis*) has two syntenic regions (paralogons) corresponding to each single region in the human genome (Jailon *et al.*, 2004). Comparative mapping, furthermore, shows that paralogons of pufferfish (*Tetraodon*), zebrafish (*Danio*) (Woods *et al.*, 2005) and medaka (*Oryzias*) (Kasahara *et al.*, 2007) are homologous. This supports the view that the FSGD occurred prior to the divergence of these teleosts.

The earliest inklings of the FSGD came from comparative analysis of *Hox* genes and clusters in different chordate lineages (Amores *et al.*, 1998, 2004; Chiu *et al.*, 2002, 2004). *Hox* genes, which encode transcription factors that play a central role in embryonic patterning of the body plan, are usually organized in clusters in the genome, although there are exceptions in some invertebrate lineages (Monteiro and Ferrier, 2006). Evidence to date suggests the basal state of *Hox* clusters in jawed vertebrates is four (A,B,C,D),

K.E.Chambers, R.McDaniell, J.D.Raincrow, M.Deshmukh, C.h.Chiu
Department of Genetics, Rutgers University, Piscataway, NJ, USA
E-mail: kchamber@wiley.com, rmmcdaniell@hotmail.com,
raincrow@biology.rutgers.edu, maya.deshmukh@gmail.com,
chiu@biology.rutgers.edu

P.F. Stadler
Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany; and
RNomics Group, Fraunhofer Institut für Zelltherapie und Immunologie — IZI Perlickstrasse 1, D-04103 Leipzig, Germany; and
Department of Theoretical Chemistry, University of Vienna, Währingerstraße 17, A-1090 Wien, Austria; and
Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA
E-mail: studla@bioinf.uni-leipzig.de

as is found in cartilaginous (shark (Chiu *et al.*, 2002; Kim *et al.*, 2000; Prohaska *et al.*, 2004; Venkatesh *et al.*, 2007)), lobe-finned (human (Krumlauf, 1994), latimeria (Koh *et al.*, 2003; Powers and Amemiya, 2004)), and basal ray-finned (bichir (Chiu *et al.*, 2004)) fishes.

In contrast, zebrafish has 7 Hox clusters that house expressed genes (**HoxAa, Ab, Ba, Bb, Ca, Cb, Da** (Amores *et al.*, 1998), where **Aa** and **Ab** duplicated clusters are each orthologous to the single **HoxA** cluster of outgroup taxa such as human (Amores *et al.*, 1998, 2004; Chiu *et al.*, 2002) Recently, the **Db** cluster (the 8th cluster) in zebrafish has been found to contain a single microRNA and no open reading frames (ORFs) (Woltering and Durston, 2006). Evidence of duplicated Hox clusters is reported for additional teleosts including pufferfishes (*Takifugu rubripes* and *Tetraodon nigroviridis* (Jaillon *et al.*, 2004; Amores *et al.*, 2004; Aparicio *et al.*, 2002), medaka (*Oryzias latipes* (Kasahara *et al.*, 2007; Kurosawa *et al.*, 2006; Naruse *et al.*, 2000), striped bass (*Morone saxatilis* (Snell *et al.*, 1999)), killifish (*Fundulus heteroclitus* (Misof and Wagner, 1996)), cichlids (*Oreochromis niloticus* (Santini and Bernardi, 2005), *Astatotilapia burtoni* (Hoegg *et al.*, 2007; Thomas-Chollier and Ledent, 2008)), salmon (*Salmo salar* (Moghadam *et al.*, 2005b; Mungpakdee *et al.*, 2008)), rainbow trout (*Oncorhynchus mykiss* (Moghadam *et al.*, 2005a)), goldfish (*Carassius auratus* (Luo *et al.*, 2007)), and Wuchang bream (*Megalobrama amblycephala* (Zou *et al.*, 2007)).

Comparative analysis of Hox clusters and genes in teleosts showed that the duplicated Hox **a** and **b** clusters have experienced divergent resolution producing variation in gene content (Lynch and Force, 2000; Prohaska and Stadler, 2004) and increased rates of substitution in both protein coding (Chiu *et al.*, 2000; Wagner *et al.*, 2005; Crow *et al.*, 2006) and noncoding (Chiu *et al.*, 2002, 2004; Tumpel *et al.*, 2006) sequences. Consistent with a shared duplication, the Hox paralogs form two distinct **a** and **b** clades (Amores *et al.*, 2004). All teleosts examined to-date represent only two species-rich actinopterygian clades, the Ostariophysii (e.g. zebrafish), and Euteleostei (Acanthopterygii: pufferfishes, killifish, medaka, bass, and cichlids; Salmoniformes: salmon, trout), comprising 6,000 and 16,000 species, respectively (Nelson, 1994) (Figure 1).

One may ask whether the FSGD is directly responsible for the biological diversification (i.e. speciosity) of ray-finned fishes (Vogel, 1998; Wittbrodt *et al.*, 1998; Meyer and Schartl, 1999; Venkatesh, 2003; Postlethwait *et al.*, 2004; Meyer and Van de Peer, 2005; Volff, 2005). Alternatively, species-richness and large-scale duplications have to be considered as independent phenomena. The examination of the actinopterygian fossil record (Donoghue and Purnell, 2005) shows that there are 11 extinct clades between teleosts and their closest living relatives. The authors conclude that the character acquisitions often attributed as synapomorphies of

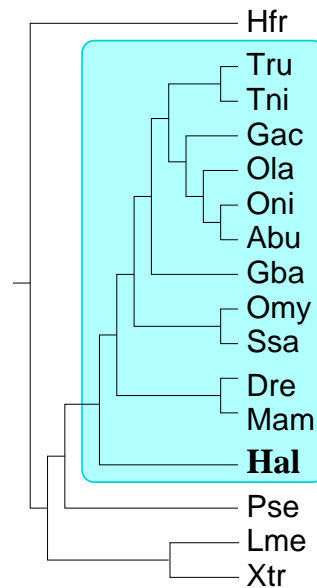


Fig. 1 Simplified phylogeny of jawed vertebrates, with focus on ray-finned fishes (actinopterygians). The jawed vertebrate clade consists of three branches, the cartilaginous (Chondrichthyes), the lobe-finned (Sarcopterygii), and ray-finned (Actinopterygii) fishes (Le *et al.*, 1993; Venkatesh *et al.*, 2001; Kikugawa *et al.*, 2004; Inoue *et al.*, 2003); the close relationship of cichlids is supported by both nuclear genes and phylogenomics data (Chen *et al.*, 2004; Steinke *et al.*, 2006).

Abbreviations: Hfr, *Heterodontus francisci* (horn shark); Xtr, *Xenopus tropicalis* (frog); Lme, *Latimeria menadoensis* (coelacanth); Pse, *Polypterus senegalus* (bichir); Hal, *Hiodon alosoides* (goldeye); Dre, *Danio rerio* (zebrafish); Mam, *Megalobrama amblycephala*; Ssa, *Salmo salar* (salmon); Omy, *Oncorhynchus mykiss* (rainbow trout); Gba, *Gonostoma bathyphilum* (lightfish); Gac, *Gasterosteus aculeatus* (three-spined stickleback); Ola, *Oryzias latipes* (medaka); Oni, *Oreochromis niloticus* (nile tilapia); Abu, *Astatotilapia burtoni*; Tni, *Tetraodon nigroviridis* (spotted green pufferfish); Tru, *Takifugu rubripes* (Japanese pufferfish)

derived teleost fishes arose gradually in ray-finned fish phylogeny with many innovations already predated the FSGD. Many of these extinct clades that have been shown to predate the FSGD were species rich themselves. Hence fossil evidence suggests that the FSGD is uncoupled to species richness. By showing that the species-poor Osteoglossomorpha exhibit duplicated Hox clusters, we add molecular evidence to this view.

Evidence from a handful of molecular evolution studies is consistent with this hypothesis. Phylogenetic analyses of four Hox genes (*HoxA11*, *HoxB5*, *HoxC11*, and *HoxD4*) (Crow *et al.*, 2006), duplicated ion and water transporter genes in eels (Cutler and Cramb, 2001), three nuclear genes (*fzd8*, *sox11*, tyrosinase (Hoegg *et al.*, 2004), the ParaHox cluster (Mulley *et al.*, 2006), and combined datasets (Hurley *et al.*, 2007) in basal, intermediate and derived actinopterygians together suggest that the FSGD is coincident with the origin of teleosts. More precisely, the data place the duplication event after the divergence of bowfin (*Amia*) and

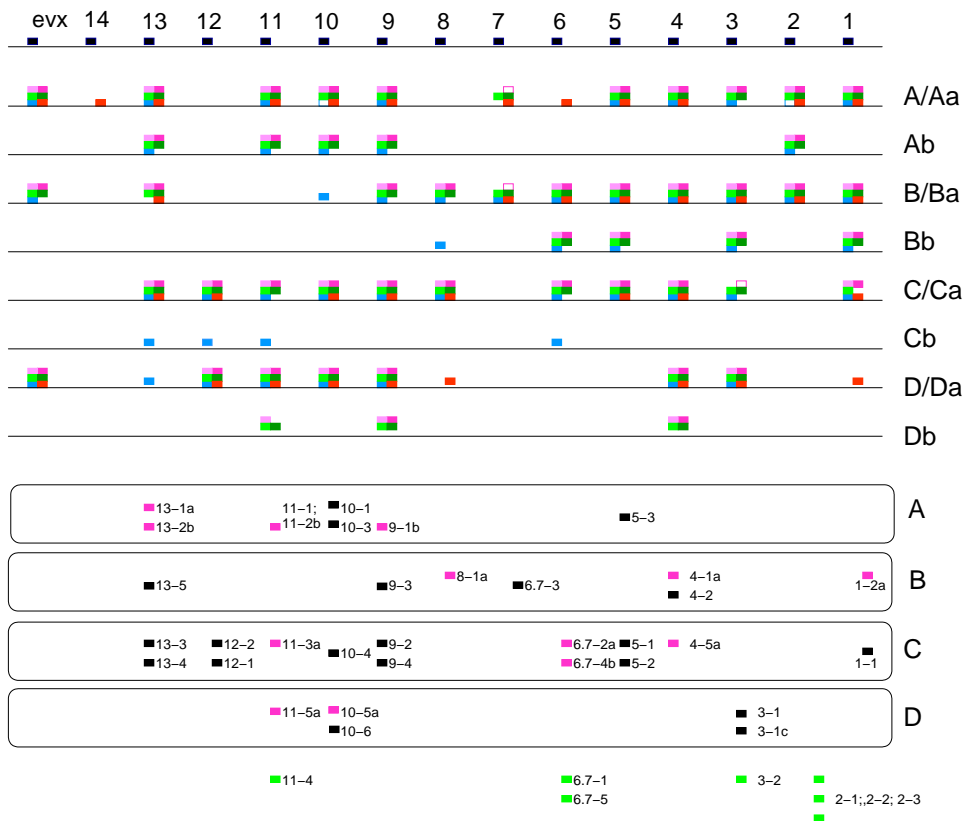


Fig. 2 *Hox* cluster complement of chordates with focus on actinopterygians. The *Hox* cluster of *Amphioxus* is shown at the top. The *Hox* genes are depicted as colored rectangles for coelacanth (outgroup; red); zebrafish (blue), medaka (light green), tilapia (dark green), Tetraodon (pink) and Fugu (magenta) are shown in the top panel. Putative goldeye *Hox* genes, as inferred from the PCR fragments, are depicted as colored rectangles in the bottom panel. Black rectangles indicate homeoboxes that are assigned to a specific paralog group and cluster (e.g. **B**) but not to a teleostean **a** or **b** clade (see text). Fuchsia rectangles indicate homeoboxes that are assigned to a specific paralog group, cluster and clade. Green rectangles depict homeobox fragments assigned to a specific paralog group but not cluster.

sturgeon but prior to the appearance ~ 135 mya of the lineages leading to 23,637 (93%) of the 23,681 extant species of present-day teleosts (Benton, 2005).

In order to assess the *Hox* complement in the earliest teleost lineages we identified *Hox* genes in the goldeye (*Hiodon alosoides*), a member of the species-poor Osteoglossomorpha (Nelson, 1994; Hurley *et al.*, 2007; Benton, 2005). Results of a PCR survey of *Hox* genes in the goldeye coupled with phylogenetic analyses of four individual *Hox* orthologs (*HoxA10*, *HoxA13-1*, *HoxA13-2*, *HoxC4*) provide conclusive evidence that the goldeye has duplicated *Hox* clusters. The organization of the goldeye *Hox* clusters, however, is significantly different from that of other teleosts, in that it has retained *Hox* genes in all eight clusters.

2 Materials and Methods

2.1 Gnathostome *Hox* Genes

Nucleotide and amino acid sequences of individual *Hox* genes analyzed in this study came from three sources: genome databases, published literature, and targeted PCR amplification using degenerate primers designed here (see below). *Amphioxus* (*Brachiostoma floridae*) homeobox sequences are from (Garcia-Fernández and Holland, 1994; Ferrier *et al.*, 2000). The representative of the cartilaginous fishes is horn shark (*Heterodontus francisci*): **HoxA** cluster, *AF479755*;

HoxD, cluster *AF224262*. The representatives of the lobe-finned fishes are coelacanth (*Latimeria menadoensis*) and frog (*Xenopus tropicalis*). Coelacanth homeobox fragments are listed in (Koh *et al.*, 2003); we (Chiu *et al.*, 2000) also sequenced the *HoxA11* ortholog (*AF287139*). Frog *Hox* clusters were taken from the Ensembl Web Browser *Xenopus tropicalis* genome JGI3: **HoxA**, scaffold29 1,777,789-2,133,531; **HoxB**, scaffold329 415,000-1,016,000; **HoxC**, scaffold280 199,492-581,365; **HoxD** scaffold353 474,676-800,000.

The representatives of the ray-finned fishes include bichir (*Polypterus senegalus*) and several teleost fishes. The bichir **HoxA** cluster was assembled from two BAC clones with accession numbers *AC126321* and *AC132195* as in (Chiu *et al.*, 2004). Zebrafish (*Danio rerio*) *Hox* clusters were assembled from PAC clones: **HoxAa**, *AC107364*; **HoxAb**, *AC107365* (with an alteration of nucleotide 79,324 from T to C to avoid a premature stop codon); **HoxBa**, *BX297395*, *AL645782*; **HoxBb**, *AL645798*; **HoxCa**, *BX465864* and *BX005254*; the **HoxCb** cluster was taken from Ensembl Web Browser *Danio rerio* genome (Zv5); **HoxDa**, *BX322661*. The zebrafish **HoxDb** cluster does not house *Hox* genes (Woltering and Durston, 2006) and was excluded in this study. Nile tilapia (*Oreochromis niloticus*) **HoxAa**, *AF533976*; striped bass (*Morone saxatilis*) **HoxAa**, *AF089743*. Medaka (*Oryzias latipes*) **Hox** clusters *AB232918-AB232924*. Spotted-green pufferfish (*Tetraodon nigroviridis*) *Hox* clusters were

extracted from the Tetraodon Genome Browser¹: **HoxAa**, chr.21. 2,878,001-3,153,406; **HoxAb**, chr.8 6,506,471-6,727,504; **HoxBa** chr.Un 37,928,410-38,293,032; **HoxBb**, chr.2 1,321,876-1,537,033; **HoxC**, chr.9 4,083,941-4,353,227; **HoxDa**, chr.2 10,975,763-11,218,409 (a T was deleted at position 11,134,740 in order to shift back to correct frame); **HoxDb**, chr.17 9,471,3559,694,740. Japanese pufferfish (*Takifugu rubripes*) *Hox* clusters were acquired from the Ensembl genome browser (assembly FUGU 2.0). The **HoxAa** cluster is constructed from the entire scaffold 47, the **HoxAb** cluster is constructed from scaffold 330, see (Chiu *et al.*, 2002). Short homeobox fragments for QM analysis were in addition taken from (Prohaska and Stadler, 2004).

2.2 PCR amplification, cloning, and sequencing

Whole genomic DNA was extracted from ~ 80 milligrams of ethanol preserved tissue of goldeye (*Hiodon alosoides*) and lightfish (*Gonostoma bathyphilum*) using the DNeasy kit (Qiagen) and protocols.

PCR amplification of an 81 base pair (bp) fragment of the highly conserved homeobox of PG1-8 was performed using a degenerate homeobox primer pair [334: 5-GAR YTI GAR AAR GAR TTY-3; 335: 5-ICK ICK RTT YTG RAA CAA-3]. PCR amplification of an 114 bp fragment of the highly conserved homeobox of PG913 was performed using the degenerate primers [*HB913Forward*: 5-AAA GGA TCC TGC AGA ARM GNT GYC CNT AYA SNA A-3; *HB113Reverse*: 5-ACA AGC TTG AAT TCA TNC KNC KRT TYT GRA ACC A-3]. PCR amplifications were performed with AmpliTaq Gold DNA polymerase (Applied Biosystems) using the following cycling parameters: initial denaturation at 95°C for 5 min, 30 cycles of 95°C for 1 min, 50°C for 1 min, and 72°C for 1 min, and final extension at 72°C for 10 min. Final concentration of MgCl₂ was 3.5 millimolar. Amplified fragments were purified by agarose gel extraction (Qiagen) and cloned into a pGEM-T Easy vector (Promega) following the manufacturers protocol. Clones containing inserts of the correct size were identified using colony PCR and sequenced at the UMDNJ-RWJMS DNA Sequencing and Synthesis Core Facility². For each clone, both strands were sequenced using T7 and SP6 sequencing primers.

2.3 Initial assignment of PCR fragments

The 81 bp and 114 bp long sequences of PG1-8 and PG9-13 homeoboxes, respectively, were compared with the corresponding sequence fragments from a range of chordates (see above). The membership of each PCR fragment to one of the

paralog groups *Hox1-Hox13* was initially determined based on nucleotide and amino acid sequence similarity to published *Hox* sequences using blast (Altschul *et al.*, 1990, 1997). The second layer of analysis used neighbor-joining (Saitou and Nei, 1987) trees with deduced amino acid sequences (see Electronic Supplement) and assigned goldeye PCR fragments based on assigned the identity of the subtree in which they are located. With the exception of the “middle-group paralogs” *Hox4-Hox7*, we find that the paralog-groups are reconstructed as monophyletic clades (with the exception of the posterior sequences from *Amphioxus* (Garcia-Fernàndez and Holland, 1994; Ferrier *et al.*, 2000).

2.4 Assignment by Quartet Mapping

All subsequent analyses were performed using homeobox nucleotide sequences. Middle-group genes were identified using Quartet Mapping (QM), see (Nieselt-Struwe and von Haeseler, 2001) and an application of QM to homeobox PCR fragments from lower vertebrates (Stadler *et al.*, 2004) for additional details. To this end, we use the teleost homeobox sequences from (Amores *et al.*, 2004), the collection of homeobox fragments from (Prohaska and Stadler, 2004), sequences of human, shark, coelacanth and the bichir **HoxA** cluster (Chiu *et al.*, 2004) as well as sequences from our own unpublished PCR study of the bichir (Raincrow *et al.*, in preparation). We first determine QM support for paralog groups PG4, PG5, and the combination of PG6 and PG7. For those sequences that are not identified as PG4 homeoboxes, we re-run the analysis computing support for PG5, PG6, and PG7.

In a second experiment we then consider trees of the form $((\{x\}, R), (U, (V, W)))$ or $((\{x\}, (R, U)), (V, W))$, where $\{x\}$ denotes the query sequence from *Hiodon* and $\{R, U, V, W\} = \{PG4, PG5, PG6, PG7\}$ are the sets of known homeobox sequences from the four middle paralog groups. Together with the query sequence, we thus consider quintets, which can be represented in the form of six inequivalent quartets depending on which pair of paralog groups form a common subtree:

$((\{x\}, R)|(U, (V, W)))$; $((\{x\}, R)|(V, (U, W)))$; $((\{x\}, R)|(W, (U, V)))$;

$((\{x\}, (R, U))|(V, W))$; $((\{x\}, (R, V))|(U, W))$; $((\{x\}, (R, W))|(U, V))$.

We analyze each of these six quartets using quartet mapping, i.e., we determine which assignment of the four paralog groups to R, U, V, W yields the maximal support for the tree. This yields a support value for each *Hiodon* query sequence x to be placed in a common subtree with either a single paralog group or with a pair of paralog groups. Ideally, x is placed together with the same paralog group R three times and placed together with the combination of R and one other paralog group in the remaining three quartets. Our implementation `quartm` of the Quartet Mapping method performs

¹ http://www.genoscope.cns.fr/externe/tetranew/entry_ggb.html

² <http://www2.umdj.edu/dnalbweb>

this quartet analysis of quintets automatically. The program can be free downloaded from the authors' website³.

2.5 Assignment by phylogenetic analysis

The quartet mapping analysis was complemented by the construction of neighbor joining (Saitou and Nei, 1987) and maximum parsimony (Swofford, 2003) trees from the same datasets. In the next step we used the same procedure separately for each paralog group to assign a sequence to one of the four gnathostome clusters **HoxA**, **HoxB**, **HoxC**, **HoxD**. In the final step we then attempted to resolve the assignment of the Hiodon PCR fragments from each class to one of the two teleost-specific paralog groups.

2.6 Sequencing of four *Hox* orthologs

All PCR amplifications were performed with AmpliTaq Gold DNA polymerase (Applied Biosystems). Cloning and sequencing were performed as described above.

Goldeye duplicated *HoxA13-1* and *HoxA13-2* sequences and the lightfish **HoxA13b**-like sequence (Figures 3a and 4) were PCR amplified using universal *HoxA13* primers sequences (Chiu *et al.*, 2004) using the following PCR conditions (initial denaturation at 95°C for 5 min, 30 cycles of 95°C for 1 min, 53°C for 1 min, and 72°C for 3 min, and final extension at 72°C for 10 min. Final concentration of MgCl₂ was 2.0 millimolar). The lightfish *Hoxa13b*-like sequence is deposited in Genbank ([bankit1122802](#)); the goldeye duplicated *HoxA13.1* and *HoxA13.2* sequences have accession numbers [bankit1122788](#) and [bankit1122792](#), respectively.

Two overlapping primer pairs were used to PCR amplify the goldeye *HoxA10*-like sequence (Figure 3c and Supplemental Figure 2). The first set of degenerate primers (*HoxA10Uforward*: 5-CDG TNC CVG GYT ACT TCC G-3; *HoxA10Ureverse*: 5-CCC AAC AAC AKR ARA CTA CC-3) amplify approximately the last third of exon 1, the intron, and most of exon 2 using the following cycling parameters (initial denaturation at 95°C for 5 min, 30 cycles of 95°C for 1 min, 55°C for 1 min, and 72°C for 1 min, and final extension at 72°C for 10 min. Final concentration of MgCl₂ was 3.5 millimolar). To amplify the N terminal portion of exon 1 we designed a forward primer (*PFCA75*: 5-TTT GYW CRA GAA ATG TCA GC-3) from an evolutionarily conserved noncoding sequence (*PFCAEF75*; Raincrow *et al.*, in preparation) immediately upstream of the *HoxA10* start codon. PCR using this forward primer and a reverse primer (*Halaxon1R*: 5-CCT TAG AAG TTG CAT AAG CC-3) that is specific to the goldeye *HoxA10*-like exon 1 sequence (described above),

was performed under the reaction conditions (initial denaturation at 95°C for 5 min, 30 cycles of 95°C for 1 min, 55°C for 1 min, and 72°C for 1 min, and final extension at 72°C for 10 min. Final concentration of MgCl₂ was 3.0 millimolar). The *HoxA10*-like sequence of goldeye built from a contig of these overlapping PCR fragments, spanning from the promoter to exon 2, is deposited in Genbank ([bankit1122799](#)).

The *HoxC4* ortholog of bichir (*Polypterus senegalus*, Pse; [bankit1123044](#), [bankit1123047](#) and the *HoxC4a*-like paralog of goldeye (Hal; Genbank [bankit1122797](#)) were amplified with a degenerate primer pair (*HoxC4Forward*: 5-CAT GAG CTC GTY TTT GAT GGA3; *HoxC4Reverse*: 5-AYT TCA TCC TKC GGT TCT GA-3) using the following PCR conditions (initial denaturation at 95°C for 5 min, 30 cycles of 95°C for 1 min, 53°C for 1 min, and 72°C for 3 min, and final extension at 72°C for 10 min. Final concentration of MgCl₂ was 2.0 millimolar).

2.7 Phylogenetic analysis of exon 1 sequences

Alignments of *Hox* gene nucleotide sequences were done using the *clustalW* algorithm (Thompson *et al.*, 1994) in the software package MacVector, version 8.1.1, using default settings. Nucleotide sequences were trimmed so each sequence was of equal length. Alignments of *Hox* gene predicted amino acid sequences were done using the *clustalW* algorithm in the software package MacVector version 8.1.1 using default settings. Amino acid alignments were corrected by eye and trimmed so each sequence was of equal length. Alignments can be viewed in the Electronic Supplement.

Maximum Parsimony trees were created using PAUP* v4.0b10 (Swofford, 2003) under the parsimony optimality criterion. Heuristic searches were performed under default settings. Neighbor-Joining (Saitou and Nei, 1987) trees were also created using the PAUP* v4.0b10 package using the distance optimality criterion with default settings. Maximum Likelihood trees were obtained using GARLI v0.951 (Zwickl, 2006). Default settings were used unless otherwise stated below. Starting trees were obtained using heuristic search under the likelihood optimality criterion in PAUP* v4.0b10 (Swofford, 2003), default settings were used. The substitution model was set to the 2 rate model which corresponds to the HKY85 model. Under the Run Termination criteria "Bootstrap repetitions" was set to 2,000 and "Generations without improving topology" was set to 5,000 as suggested in the GARLI manual when using bootstrap repetitions. For all three methods, node confidence was scored using the bootstrap resampling method and 50% cutoff.

Bayesian trees were obtained using MrBayes v3.1.2 (Ronquist and Huelsenbeck, 2003) and the parallel version of MrBayes v3.1.2 (Altekar *et al.*, 2004). MrBayes settings were as follows: 2 rate substitution model, relative rate distribution = gamma, number of generations = 1,000,000, sam-

³ <http://www.bioinf.uni-leipzig.de/Software/quartm/>

ple freq = 1,000, number of chains = 4, and temperature = 0.2. "Burn-in" was assessed using the "sump" command. Normally, the first 1 or 2 trees were discarded as "burn-in" before creating the final consensus tree. Node confidence was scored using the Bayesian posterior probability provided by the program.

Phylogenetic networks were computed using the neighbor-net algorithm (Bryant and Moulton, 2004) implemented in the SplitsTree package (Huson and Bryant, 2006) using the same distance matrices that also underlie the neighbor-joining trees.

3 Results

The first step of this study is to estimate the number of *Hox* clusters in the goldeye (*Hiodon alosoides*). Using degenerate primers that target homeoboxes (see Methods), we cloned and sequenced a total of 421 *Hox* fragments (81 and 114 bp long, depending on the primer set utilized) and 23 non-*Hox* fragments (not further analyzed). Using a combination of blast (Altschul *et al.*, 1990, 1997), similarity, Quartet Mapping (QM; (Nieselt-Struwe and von Haeseler, 2001), and phylogenetic analyses (Electronic Supplement⁴, the 421 *Hox* sequences group into 41 unique sequences (Figure 2). For each sequence, allelic exclusion tests were performed as described in (Misof and Wagner, 1996). The 41 homeobox sequences of goldeye found in this study have been deposited in GenBank *FJ015270-FJ015310*. A full list is provided in the Electronic Supplement.

As shown in Figure 2 (bottom panel), the goldeye has duplicated paralogs on each of the four *Hox* clusters. For **HoxA**-like clusters, there is evidence for duplicated group 10, 11, and 13 paralogs; **HoxB**-like clusters, group 4; **HoxC**-like clusters, groups 5, 6, 9, 12, 13; and **HoxD**-like clusters, groups 3 and 10. Strikingly, the goldeye is the only teleost fish examined to date that has evidence for retained *Hox* genes on each of the eight *Hox* clusters (**Aa**, **Ab**, **Ba**, **Bb**, **Ca**, **Cb**, **Da**, **Db**).

Phylogenetic analysis and QM mapping, however, assigned only thirteen sequences to **a** or **b** paralog clades observed in advanced teleost fishes (Figure 2). About the same number of sequences is preferentially classified with the unduplicated genes in bichir, shark, or sarcopterygians. The PCR fragments therefore do not provide enough information to decide whether the goldeye shares the *Hox* duplication with the crown teleosts, i.e., whether its eight *Hox* clusters are orthologous to the eight teleost *Hox* loci, or whether an independent duplication event occurred in Osteoglossomorpha.

Because the homeobox sequence amplified in a genomic PCR survey is so short, we chose to further investigate this

problem by examining exon sequences of four *Hox* orthologs, *HoxA13* (two paralogs), *HoxA10* and *HoxC4*. For the *HoxA13* locus, we cloned and sequenced the gene proper region of two *HoxA13*-like paralogs (*Hal13.1* and *Hal13.2*) including the beginning of exon 1 (12aa from the start codon), intron, and most of exon 2 including the homeobox. Notably, the homeodomain sequences of *Hal13.1* and *Hal13.2* are identical to homeobox fragments 13.1 and 13.2, respectively, isolated in our independent PCR survey of *H. alosoides* whole genomic DNA.

Interestingly, while homeobox fragments *13.1* and *13.2* are tentatively assigned as *HoxA13a* and *HoxA13b* (Figure 2), gene tree reconstructions using *Hal13.1* and *Hal13.2* exon 1 amino acid sequences (Figure 3a) show that both *HoxA13*-like paralogs of goldeye do not group in either the *HoxA13a* or *HoxA13b* clades of teleost fishes. Instead, both *HoxA13* paralogs of goldeye branch at the base of teleosts, prior to the duplication but after divergence of bichir (*P. senegalus*), the most basal living lineage (Chiu *et al.*, 2004; Mulley *et al.*, 2006). Gene trees reconstructed using exon 1 nucleotide sequences do not resolve the phylogenetic position of the two *HoxA13*-like paralogs (see also Supplemental Figure 1a).

We examined the exon 1 nucleotide sequences of each *HoxA13*-like paralog in goldeye and did not detect evidence for gene conversion (data not shown). Interestingly though, when we examined the predicted primary amino acid sequence of *Hal13.1* and *Hal13.2* paralogs, we found that they share many amino acids at positions that have diverged in the duplicated paralogs of all crown teleosts (zebrafish (Chiu *et al.*, 2002), medaka (Kasahara *et al.*, 2007; Naruse *et al.*, 2000; Kurosawa *et al.*, 2006), tilapia (Santini and Bernardi, 2005), lightfish (this study) and pufferfishes (Jaillon *et al.*, 2004; Aparicio *et al.*, 2002)), see Fig. 4. The amino acid positions shared by the duplicated *HoxA13*-like paralogs in goldeye are the ancestral sites, as determined by their shared presence in the bichir (*Polypterus senegalus*), which has a single **HoxA** cluster (Chiu *et al.*, 2004). We examined whether there is selection acting on synonymous substitutions (Ks) at these two loci in the goldeye (Yang, 1997), but we did not find any statistical support (data not shown). Our findings for the goldeye *HoxA13*-like paralogs are striking because they do not exhibit a pattern of sequence evolution consistent with intensive diversifying selection (van de Peer *et al.*, 2001; Crow *et al.*, 2006) following duplication. The goldeye thus may be a good model to test the predictions of the DDC model (Force *et al.*, 1999), whereby amino acid sequence divergence of duplicated paralogs may be small but divergence in regulatory sequences is large.

Using overlapping primer sets (see below), we cloned and sequenced the gene proper region of a *HoxA10*-like sequence (Figure 3b) including a promoter sequence (not shown). The homeodomain sequence of the *HoxA10*-like ortholog is an exact match to fragment 10-1 (Figure 2), assigned as a

⁴ <http://www.bioinf.uni-leipzig.de/Publications/SUPPLEMENTS/Hiodon/>

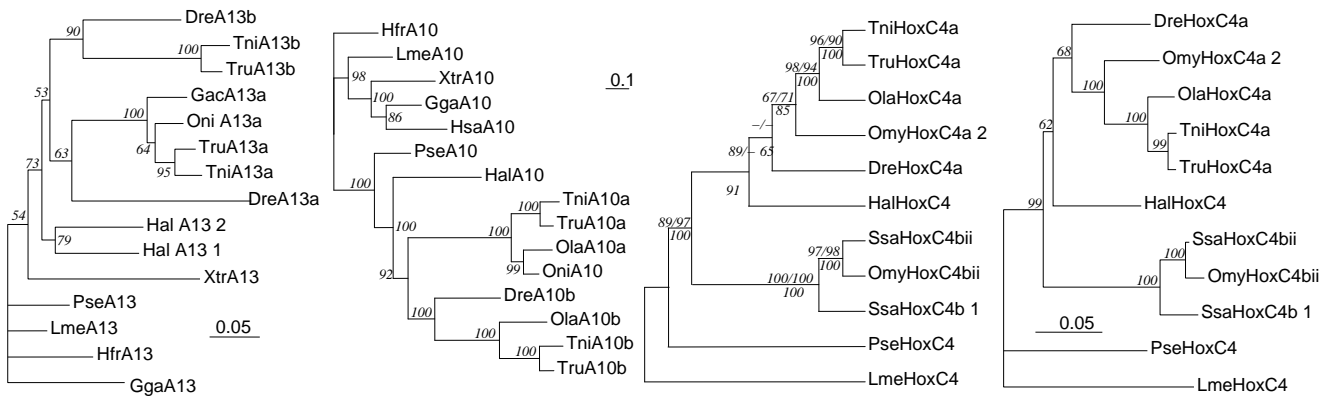


Fig. 3 Examples of phylogenetic analysis of *Hox* exon 1 sequences. Species abbreviations as in Fig. 1. (A) *HoxA13* tree reconstructed using neighbour-joining (Saitou and Nei, 1987) analysis of *HoxA13* amino acid sequences. Bootstrap support (2000 replications) are shown at the nodes. (B) *HoxA10* tree reconstructed using Bayesian (Ronquist and Huelsenbeck, 2003; Altekar *et al.*, 2004) analysis of amino acid sequences. Node confidence values of 1,000,000 generations are shown. (C) Consensus *HoxC4* tree reconstructed using Neighbor joining (Saitou and Nei, 1987), heuristic maximum parsimony (Swofford, 2003), and maximum likelihood Swofford:03,Zwickl:06 analyses of amino acid sequences. Node confidence values are listed as NJ/HMP/B. (D) Consensus *HoxC4* tree reconstructed using Neighbor joining analysis of nucleotide sequences. Node confidence values are listed as NJ/MP/B/ML. See text for details of phylogenetic analysis.

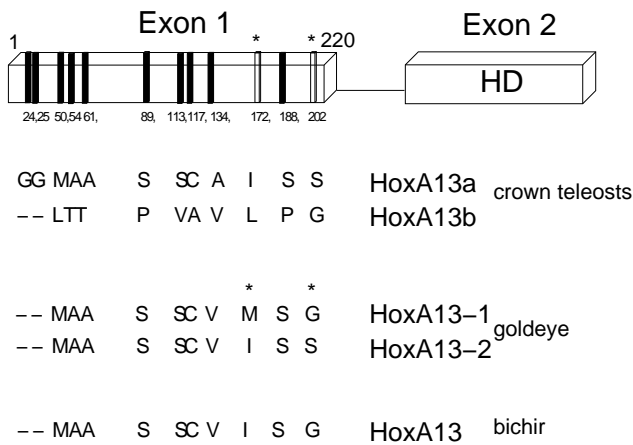


Fig. 4 Goldeye duplicated *HoxA13*-like paralogs do not diverge at the amino acid level. Cartoon depiction of *HoxA13* exon 1 and exon 2 domains. Amino acid numbers according to *HoxA13a* of pufferfish (*Takifugu*), see text. Amino acid positions (black bars) that diverge in the duplicated *HoxA13a* and *HoxA13b* paralogs of species-rich teleosts are shown and contrasted with the duplicated *HoxA13*-like paralogs of goldeye. Only two of amino acid positions diverge in goldeye (asterisks). See text for further description.

HoxA10 homeobox. As illustrated in phylogenetic analysis of exon 1 amino acid sequences, the *HoxA10*-like sequence of goldeye branches outside the duplicated *HoxA10a* and *HoxA10b* clades (Figure 3b), similarly to the *HoxA13*-like paralogs (Figure 3a). The topology of this gene tree is similar to that reported in (Hurley *et al.*, 2007) for other nuclear genes. Interestingly, the promoter of the goldeye *HoxA10*-like ortholog also has not acquired diagnostic teleostean paralog **a** and **b** specific nucleotides (not shown). There are at least two possibilities that could account for these results. First, following *Hox* cluster duplication, goldeye re-

tains only a single *HoxA10* locus that did not accumulate substitutions at an increased rate observed when both duplicated paralogs are retained following duplication in teleost crown groups (Chiu *et al.*, 2000; Wagner *et al.*, 2005; van de Peer *et al.*, 2001). In fact, phylogenetic analysis of exon 1 of the single *HoxA10b* locus in zebrafish provides strong support for branching within the teleostean **b** clade only at the amino acid (Figure 3b), but not nucleotide sequence (Supplemental Figure 1b) level. Hence, following a duplication, if one of the paralogs is immediately lost, the rate of nucleotide substitution of the remaining singlet may be conservative. A second possibility raised by our findings is that goldeye experienced a duplication that is independent from that in the crown group of ostariphsians and acanthomorphs. A third scenario, although not tenable with available data, is that goldeye experienced massive gene loss shortly after the FSGD and subsequently experienced lineage specific duplications of all or parts of its genome, including the *Hox* clusters, minimally the *HoxA*-like cluster.

Intriguingly, phylogenetic analysis of the majority of exon 1 of a *HoxC4*-like sequence found in this study provides strong support that this locus is *HoxC4a*-like at the level of amino acid (Figure 3c) and nucleotide (Figure 3d) sequences. Hence, this result supports that goldeye shares the FSGD. Importantly, the homeobox sequence of this *HoxC4a*-like locus is an identical match to our PCR homeobox survey fragment 4-5 (Figure 2) that we independently assigned as *HoxC4a* using phylogenetic methods and QM (Table 1 in the Electronic Supplement). This result, i.e., that goldeye experienced the FSGD, is consistent with the phylogenetic branching arrangement of three *Hox* genes *HoxA11α*, *HoxA11β*, and *HoxB5β* in goldeye into *HoxA11a*, *HoxA11b*, and *HoxB5b* teleostean clades, respectively (Crow *et al.*,

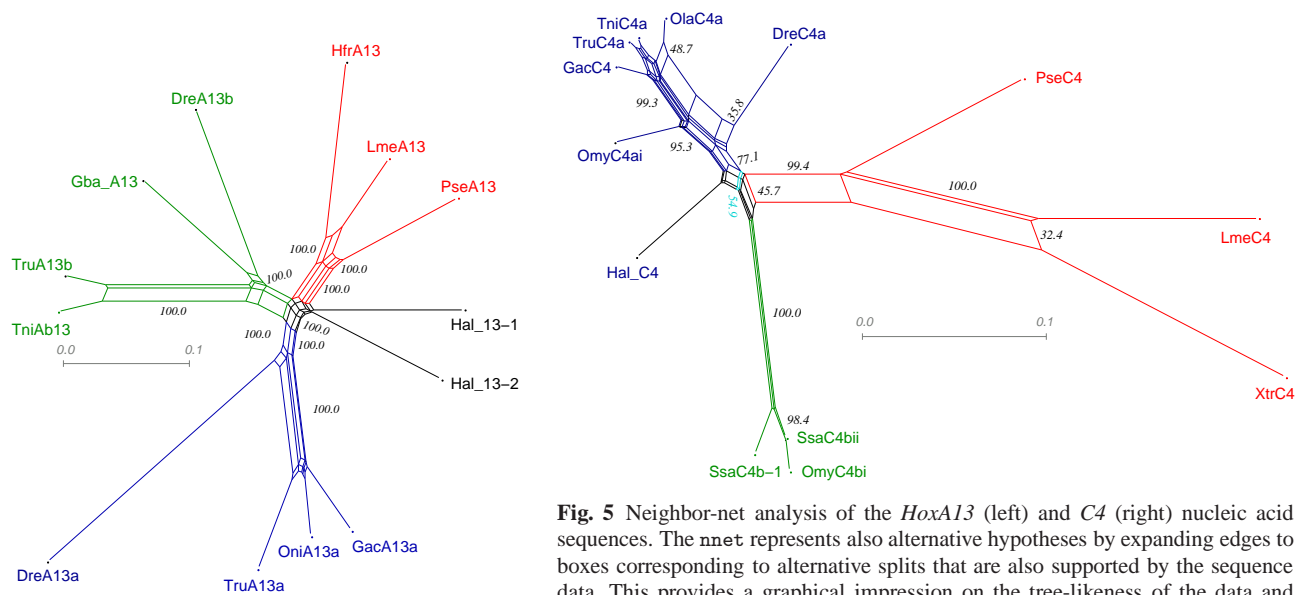


Fig. 5 Neighbor-net analysis of the *HoxA13* (left) and *C4* (right) nucleic acid sequences. The *nn*et represents also alternative hypotheses by expanding edges to boxes corresponding to alternative splits that are also supported by the sequence data. This provides a graphical impression on the tree-likeness of the data and vizualizes the signal to noise ratio of the data set.

2006). Interestingly, our PCR survey above detected two unique *HoxA11*-like homeobox fragments (*11-1*, *11-2*, Figure 2) that both are assigned, with weak support, to be *HoxA11b*-like. Our PCR screen did not yield *HoxB5*-like homeobox sequences.

4 Discussion

Our findings contribute to the understanding of the *Hox* complement in a basal teleost lineage (Figure 2) and permit inferences on when duplicate *Hox* paralogs have been lost in actinopterygian phylogeny.

While acantomorpha have completely lost one of the **HoxC** duplicates, and ostariophysys as well as Salmoniformes have lost all protein coding genes from one of the **HoxC** duplicates, goldeye has retained *Hox* genes of all eight clusters. As illustrated in Figure 2, goldeye in particular possesses duplicate paralogs of *HoxB4*, *HoxC5*, *HoxC6*, *HoxD3*, and *HoxD10*. In contrast zebrafish, with the exception of *HoxC6* (Amores *et al.*, 1998), medaka (Kasahara *et al.*, 2007; Naruse *et al.*, 2000; Kurosawa *et al.*, 2006) cichlids (Santini and Bernardi, 2005; Hoegg *et al.*, 2007; Thomas-Chollier and Ledent, 2008), and pufferfishes (Aparicio *et al.*, 2002; Jallion *et al.*, 2004), each possess at most a single copy of these loci (Figure 2). Based on fossil evidence, we infer that these genes were lost in the time interval spanning from 250 million years ago (*Amia*) to 135 million years ago (appearance of ostariophysans) (Benton, 2005).

The functional consequences of this seeming bias in gene losses remain to be explored. One prediction is that the remaining single ortholog of each locus may exhibit a pattern of sequence evolution diagnostic of negative or stabilizing

selection, which is in contrast to the pattern of strong positive selection (i.e. molecular adaptation with $K_a/K_s > 1$) that has been reported when duplicated paralogs are retained, such as the zebrafish *HoxC6a* and *HoxC6b* paralogs (van de Peer *et al.*, 2001), **HoxA** cluster duplicated paralogs of ostariophysan and acanthomorph lineages (Chiu *et al.*, 2000; Wagner *et al.*, 2005) and other nuclear loci (Brunet *et al.*, 2006).

The duplication of the *Hox* gene system in goldeye together with previously reported duplications (relative to the gnathostome ancestor) of several other nuclear genes in other bony tongues (Hoegg *et al.*, 2004) suggests that we are dealing with a whole-genome duplication. A genome duplication, or the possession of a duplicated *Hox* system in particular, is therefore uncoupled from species-richness. Our results emphasize the genome plasticity of actinopterygians in general and suggest that different mechanisms may be at work in the earliest (species poor) versus later (species rich) teleost fishes.

Strictly speaking, our data fail to conclusively resolve the question whether or not the duplicated *Hox* clusters in goldeye are true orthologs of the eight teleostean clusters. As illustrated in Figure 3a, the branch length of each *HoxA13*-like sequence in goldeye is long, suggesting they derive from an ancient duplication and not a lineage specific duplication as observed in paddlefish for *HoxB5* duplicated paralogs (Crow *et al.*, 2006). The ambiguity of the phylogenetic analysis, furthermore, in itself implies that the duplication observed in osteoglossomorpha must have been very *close* in time to the divergence of this lineage from crown teleosts, a conclusion also drawn in (Crow *et al.*, 2006). This is illustrated nicely by the phylogenetic networks in Figure 5,

which show that the phylogenetic signal (branch lengths) separating the FSGD from the divergence of Osteoglossomorpha and crown teleosts is comparable to the noise inherent in the available data.

In conclusion, our analysis is consistent both with independent duplications in both lineages shortly after the osteoglossomorpha-crown teleost split, and with the — more parsimonious — interpretation of a single FSGD pre-dating this divergence (Crow *et al.*, 2006). We suspect that a definitive resolution of this question will require genome-wide data as well as a denser taxon sampling at key points in actinopterygian phylogeny.

Acknowledgements Goldeye and lightfish genomic DNAs are a gift from Dr. Guillermo Orti. We are grateful to Ms. Inna Zamanskaya and Mr. Suley Kuyumcu for technical assistance, and Bärbel M. R. Stadler for carefully proofreading this manuscript. This work was supported by the National Science Foundation, MCB 0447478 (C.H.C.) and the Bioinformatics Initiative of the Deutsche Forschungsgemeinschaft, BIZ-6/1-2 (P.F.S).

Supplemental Material

An extensive **Electronic Supplement** provides further details on the phylogenetic analysis of the exon-1 sequences (Supplementary Figure 1), Tables detailing the phylogenetic and quartet mapping analysis of the PCR fragments, as well as machine readable files containing sequences and alignments. The electronic Supplement can be found at <http://www.bioinf.uni-leipzig.de/Publications/SUPPLEMENTS/Hiodon/>.

References

- Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F, 2004. Parallel Metropolis-coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* 20:407–415.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410.
- 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 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 YL, Pomeroy J, Singer A, Amemiya C, Postlethwait J, 2004. Developmental roles of pufferfish *Hox* clusters and genome evolution in ray-finned 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, Verhoeve F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJK, Dogget N, Zharkikh A, Tavtigian SV, Pruss D, Barstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, H. TY, 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.
- Benton MJ, 2005. *Vertebrate Paleontology*. Malden: Blackwell, 3rd edn.
- Brunet FG, Crollius HR, Paris M, Aury JM, Gibert P, Jailion O, Laudet V, Robinson-Rechavi M, 2006. Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. *Mol Biol Evol* 23:1808–1816.
- Bryant D, Moulton V, 2004. Neighbor-net: An agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* 21:255–265.
- Chen WJ, Orti G, Meyer A, 2004. Novel evolutionary relationship among four fish model systems. *Trends Genet* 20:424–431.
- Chiu CH, Amemiya C, Dewar K, Kim CB, Ruddle FH, Wagner GP, 2002. Molecular evolution of the *HoxA* cluster in the three major gnathostome lineages. *Proc Natl Acad Sci USA* 99:5492–5497.
- Chiu CH, Dewar K, Wagner GP, Takahashi K, Ruddle F, Ledje C, Bartsch P, Scemama JL, Stellwag E, Fried C, Prohaska SJ, Stadler PF, Amemiya CT, 2004. Bichir *HoxA* cluster sequence reveals surprising trends in ray-finned fish genomic evolution. *Genome Res* (p. 14). 11–17.
- Chiu CH, Nonaka D, Xue L, Amemiya CT, Wagner GP, 2000. Evolution of *HoxA11* in lineages phylogenetically positioned along the fin-limb transition. *Mol Phylogenet Evol* 17:305–316.
- Christoffels A, Koh EGL, Chia JM, Brenner S, Aparicio S, Venkatesh B, 2004. *Fugu* genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol Biol Evol* 21:1146–1151.
- Crow KD, Stadler PF, Lynch VJ, Amemiya C, Wagner GP, 2006. The fish-specific *hox* cluster duplication is coincident with the origin of teleosts. *Mol Biol Evol* 23:121–136.
- Cutler CP, Cramb G, 2001. Molecular physiology of osmoregulation in eels and other teleosts: the role of transporter isoforms and gene duplication. *Comp Biochem Physiology A* 130:551–564.
- Donoghue PCJ, Purnell MA, 2005. Genome duplication, extinction, and vertebrate evolution. *Trends Ecol Evol* 20:312–319.

- Ferrier DEK, Minguillón C, Holland PWH, Garcia-Fernández J, 2000. The amphioxus *Hox* cluster: deuterostome posterior flexibility and *Hox14*. *Evol Dev* 2:284–293.
- Force A, Lynch M, Pickett FB, Amores A, Yan YI, Postlethwait J, 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545.
- Garcia-Fernández J, Holland PW, 1994. Archetypal organization of the amphioxus *hox* gene cluster. *Nature* 370:563–566.
- Hoegg S, Boore JL, Kuehl JV, Meyer A, 2007. Comparative phylogenomic analyses of teleost fish *Hox* gene clusters: lessons from the cichlid fish *Astatotilapia burtoni*. *BMC Genomics* 8:317.
- Hoegg S, Brinkmann H, Taylor JS, Meyer A, 2004. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *J Mol Evol* 59:190–203.
- Hurley IA, Mueller RL, Dunn KA, Schmidt EJ, Friedman M, Ho RK, Prince VE, Yang Z, Thomas MG, , Coates MI, 2007. A new time-scale for ray-finned fish evolution. *Proc Biol Sci* 274:489–498.
- Huson DH, Bryant D, 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23:254–267.
- Inoue JG, Miya M, Tsukamoto K, Nishida M, 2003. Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the “ancient fish”. *Mol Phylog Evol* 26:110–120.
- Jaillon O, Aury J, Brunet F, Petit J, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S, Jaffe D, Fisher S, Lutfalla G, Dossat C, Segurens B, Dasilva C, Salanoubat M, Levy M, Boudet N, Castellano S, Anthouard V, Jubin C, Castelli V, Katinka M, Vacherie B, Biémont C, Skalli Z, Caticolico L, Poulain J, De Berardinis V, Cruaud C, Duprat S, Brottier P, Coutanceau J, Gouzy J, Parra G, Lardier G, Chapple C, McKernan K, McEwan P, Bosak S, Kellis M, Volff J, Guigó R, Zody M, Mesirov J, Lindblad-Toh K, Birren B, Nusbaum C, Kahn D, Robinson-Rechavi M, Laudet V, Schachter V, Quétier F, Saurin W, Scarpelli C, Wincker P, Lander E, Weissenbach J, Roest Crollius H, 2004. Genome duplication in the teleost fish tetraodon *nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431:946–957.
- Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y, Jindo T, Kobayashi D, Shimada A, Toyoda A, Kuroki Y, Fujiyama A, Sasaki T, Shimizu A, Asakawa S, Shimizu N, Hashimoto S, Yang J, Lee Y, Matsushima K, Sugano S, Sakaizumi M, Narita T, Ohishi K, Haga S, Ohta F, Nomoto H, Nogata K, Morishita T, Endo T, Shin-I T, Takeda H, Morishita S, Kohara Y, 2007. The medaka draft genome and insights into vertebrate genome evolution. *Nature* 447:714–719.
- Kikugawa K, Katoh K, Kuraku S, Sakurai H, Ishida O, Iwabe N, Miyata T, 2004. Basal jawed vertebrate phylogeny inferred from multiple nuclear DNA-coded genes. *BMC Biol* 2:3.
- Kim CB, Amemiya C, Bailey W, Kawasaki K, Mezey J, Miller W, Minosima S, Shimizu N, P. WG, Ruddle F, 2000. *Hox* cluster genomics in the horn shark, *heterodontus francisci*. *Proc Natl Acad Sci USA* 97:1655–1660.
- Koh EGL, Lam K, Christoffels A, Erdmann MV, Brenner S, Venkatesh B, 2003. *Hox* gene clusters in the Indonesian coelacanth, *Latimeria menadoensis*. *Proc Natl Acad Sci USA* 100:1084–1088.
- Krumlauf R, 1994. *Hox* genes in vertebrate development. *Cell* 78:191–201.
- Kurosawa G, Takamatsu N, Takahashi M, Sumitomo M, Sanaka E, Yamada K, Nishii K, Matsuda M, Asakawa S, Ishiguro H, Miura K, Kurosawa Y, Shimizu N, Kohara Y, Hori H, 2006. Organization and structure of *hox* gene loci in medaka genome and comparison with those of pufferfish and zebrafish genomes. *Gene* 370:75–82.
- 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.
- Luo J, Stadler PF, He S, Meyer A, 2007. PCR survey of *Hox* genes in the Goldfish *Carassius auratus auratus*. *J Exp Zool B Mol Devel Evol* 308B:250–258.
- Lynch M, Force A, 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154:459–473.
- 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, 2005. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *BioEssays* 27:937–945.
- Misof BY, Wagner GP, 1996. Evidence for four *Hox* clusters in the killifish *Fundulus Heteroclitus* (teleostei). *Mol Phylog Evol* 5:309–322.
- Moghadam HK, Ferguson MM, Danzmann RG, 2005a. Evidence for *Hox* gene duplication in rainbow trout (*Oncorhynchus mykiss*): A tetraploid model species. *J Mol Evol* 61:804–818.
- Moghadam HK, Ferguson MM, Danzmann RG, 2005b. Evolution of *Hox* clusters in salmonidae: A comparative analysis between Atlantic salmon (*Salmo salar*) and rainbow trout (*Oncorhynchus mykiss*). *J Mol Evol* 61:636–649.

- Monteiro AS, Ferrier DEK, 2006. Hox genes are not always colinear. *Int J Biol Sci* 2:95–103.
- Mulley JF, Chiu CH, Holland PW, 2006. Breakup of a homeobox cluster after genome duplication in teleosts. *Proc Natl Acad Sci USA* 103:10369–10372.
- Mungpakdee S, Seo HC, Angotzi AR, Dong X, Akalin A, Chourrout D, 2008. Differential evolution of the 13 Atlantic salmon *Hox* clusters. *Mol Biol Evol* 25:1333–1343.
- 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.
- Nelson JS, 1994. *Fishes of the world*. New York: John Wiley & Sons Inc., 3rd edn.
- Nieselt-Struwe K, von Haeseler A, 2001. Quartet-mapping, a generalization of the likelihood mapping procedure. *Mol Biol Evol* 18:1204–1219.
- Ohno S, 1970. *Evolution by gene duplication*. New York: Springer-Verlag.
- Postlethwait J, Amores A, Cresko W, Singer A, Yan YL, 2004. Subfunction partitioning, the teleost radiation and the annotation of the human genome. *Trends Genet* 20:481–490.
- Powers TP, Amemiya CT, 2004. Evidence for a hox14 paralog group in vertebrates. *Current Biol* 14:R183–R184.
- Prohaska SJ, Fried C, Amemiya CT, Ruddle FH, Wagner GP, Stadler PF, 2004. The shark HoxN cluster is homologous to the human HoxD cluster. *J Mol Evol* (p. 58). 212–217.
- Prohaska SJ, Stadler PF, 2004. The duplication of the hox gene clusters in teleost fishes. *Th Biosci* 123:89–110.
- Ronquist F, Huelsenbeck JP, 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
- Saitou N, Nei M, 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425.
- Santini S, Bernardi G, 2005. Organization and base composition of tilapia *Hox* genes: implications for the evolution of *Hox* clusters in fish. *Gene* 346:51–61.
- Sidow A, 1996. Gen(om)e duplications in the genomes of early vertebrates. *Curr Opin Genet Dev* 6:715–722.
- Snell EA, Scemama JL, Stellwag EJ, 1999. Genomic organization of the hoxa4-hoxa10 region from *Morone saxatilis*: implications for hox gene evolution among vertebrates. *J Exp Zool Mol Dev Evol* 285:41–49.
- Stadler PF, Fried C, Prohaska SJ, Bailey WJ, Misof BY, Ruddle FH, Wagner GP, 2004. Evidence for independent *Hox* gene duplications in the hagfish lineage: A PCR-based gene inventory of *Eptatretus stoutii*. *Mol Phylog Evol* 32:686–692.
- Steinke D, Salzburger W, Meyer A, 2006. Novel relationships among ten fish model species revealed based on a phylogenomic analysis using ESTs. *J Mol Evol* 62:772–784.
- Swofford DL, 2003. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods) Version 4. Sunderland, MA: Sinauer Associates. Handbook and Software.
- 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.
- Taylor JS, Van de Peer Y, Meyer A, 2001. Revisiting recent challenges to the ancient fish-specific genome duplication hypothesis. *Curr Biol* 11:R1005–R1008.
- Thomas-Chollier M, Ledent V, 2008. Comparative phylogenomic analyses of teleost fish *Hox* gene clusters: lessons from the cichlid fish *Astatotilapia burtoni*: comment. *BMC Genomics* 9:35.
- Thompson JD, Higgs DG, Gibson TJ, 1994. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties, and weight matrix choice. *Nucl Acids Res* 22:4673–4680.
- Tumpel S, Cambronero F, Wiedemann LM, Krumlauf R, 2006. Evolution of *cis* elements in the differential expression of two Hoxa2 coparalogous genes in pufferfish (*Takifugu rubripes*). *Proc Natl Acad Sci USA* 103:5419–5424.
- van de Peer Y, Taylor JS, I. B, Meyer A, 2001. The ghost of selection past: rates of evolution and functional divergence of anciently duplicated genes. *J Mol Evol* 53:436–446.
- Vandepoele K, De Vos W, Taylor JS, Meyer A, Van de Peer Y, 2004. Major events in the genome evolution of vertebrates: Paraneome age and size differ considerably between ray-finned fishes and land vertebrates. *Proc Natl Acad Sci USA* 101:1638–1643.
- Venkatesh B, 2003. Evolution and diversity of fish genomes. *Curr Opin Genet Dev* 13:588–592.
- Venkatesh B, Erdmann MV, Brenner S, 2001. Molecular synapomorphies resolve evolutionary relationships of extant jawed vertebrates. *Proc Natl Acad Sci USA* 98:11382–11387.
- Venkatesh B, Kirkness EF, Loh YH, Halpern AL, Lee AP, Johnson J, Dandona N, Viswanathan LD, Tay A, Venter JC, Strausberg RL, Brenner S, 2007. Survey sequencing and comparative analysis of the elephant shark (*Callorhynchus milii*) genome. *PLoS Biol* 5:e101.
- Vogel G, 1998. Doubled genes may explain fish diversity. *Science* 281:1119–1121.
- Volff JN, 2005. Genome evolution and biodiversity in teleost fish. *Heredity* 94:280–294.
- Wagner GP, Takahashi K, Lynch V, Prohaska SJ, Fried C, Stadler PF, Amemiya CT, 2005. Molecular evo-

- lution of duplicated ray finned fish *hoxa* clusters: Increased synonymous substitution rate and asymmetrical co-divergence of coding and non-coding sequences. *J Mol Evol* (pp. 665–676).
- Wittbrodt J, Meyer A, Schartl M, 1998. More genes in fish? *Bioessays* 20:511–512.
- Woltering JM, Durston AJ, 2006. The zebrafish *hoxDb* cluster has been reduced to a single microRNA. *Nat Genet* 38:601–602.
- Woods IG, Wilson C, Friedlander B, Chang P, Reyes DK, Nix R, Kelly PD, Chu F, Postlethwait JH, Talbot WS, 2005. The zebrafish gene map defines ancestral vertebrate chromosomes. *Genome Res* 15:1307–1314.
- Yang Z, 1997. PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555–556.
- Zou SM, Jiang XY, He ZZ, Yuan J, Yuan XN, Li SF, 2007. *Hox* gene clusters in blunt snout bream, *Megalobrama amblycephala* and comparison with those of zebrafish, fugu and medaka genomes. *Gene* 400:60–70.
- Zwickl DJ, 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. thesis, The University of Texas at Austin.