JOURNAL OF EXPERIMENTAL ZOOLOGY (MOL DEV EVOL) 308B:250–258 (2007)

PCR Survey of Hox Genes in the Goldfish *Carassius auratus auratus*

JING LUO[1,2], PETER F. STADLER[2–5], SHUNPING HE[2], AND AXEL MEYER[1]*
[1]*Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, 78457, Konstanz, Germany*
[2]*Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, Hubei 430072, P.R. China*
[3]*Lehrstuhl für Bioinformatik, Institut für Informatik, Universität Leipzig, D-04103 Leipzig, Germany*
[4]*Institut für Theoretische Chemie und Molekulare Strukturbiologie, Universität Wien, A-1090 Wien, Austria*
[5]*Santa Fe Institute, Santa Fe, New Mexico 87501*

ABSTRACT A tetraploidization event took place in the cyprinid lineage leading to goldfishes about 15 million years ago. A PCR survey for Hox genes in the goldfish *Carassius auratus auratus* (Actinopterygii: Cyprinidae) was performed to assess the consequences of this genome duplication. Not surprisingly, the genomic organization of the Hox gene clusters of goldfish is similar to that of the closely related zebrafish (*Danio rerio*). However, the goldfish exhibits a much larger number of recent pseudogenes, which are characterized by indels. These findings are consistent with the hypothesis that dosage effects cause selection pressure to rapidly silence crucial developmental regulators after a tetraploidization event. *J. Exp. Zool. (Mol. Dev. Evol.) 308B:250–258, 2007.*

How to cite this article: Luo J, Stadler PF, He S, Meyer A. 2007. PCR survey of Hox genes in the goldfish *carassius auratus auratus*. J. Exp. Zool. (Mol. Dev. Evol.) 308B:250–258.

The duplications of genes and entire genomes are thought to be important mechanisms underlying the genesis of morphological variation and functional innovation (Ohno and Atkin, '66; Ohno et al., '67; Taylor et al., 2001; Wagner, 2001; Wagner et al., 2005; Crow et al., 2006). Extensive comparative genomics studies have corroborated the proposition that gnathostomes have experienced two rounds of genome duplication (2R) (Sidow, '96; Amores et al., '98; Taylor et al., 2003; Vandepoele et al., 2004; Dehal and Boore, 2005). Moreover, another round (3R) genome duplication, the so-called ''fish-specific genome duplication'' (FSGD) took place about 320 mya ago in the ancestral lineage of the teleosts (see e.g., Amores et al., '98; Wittbrodt et al., '98; Gregory and Hebert, '99; Taylor et al., 2003; Hoegg et al., 2004; Vandepoele et al., 2004; Hoegg and Meyer, 2005; Meyer and Van de Peer, 2005; Yan et al., 2005; Crow et al., 2006).

Fishes exhibit remarkable variation in morphological, behavioral, and physiological adaptations. Teleosts comprise more than 97% of the approximately 25,000 species of actinopterygians (ray-finned fishes) (Nelson, '94) and are the most species-rich and diverse group of vertebrates. Several authors suggested that the FSGD is at least partially responsible for the species diversity of teleosts (e.g., Amores et al., '98; Wittbrodt et al., '98; Meyer and Van de Peer, 2005; Yan et al., 2005; but see Donoghue and Purnell, 2005).

Hox genes encode for homeodomain-containing transcription factors which are orthologous to the genes in the Drosophila homeotic gene clusters (McGinnis and Krumlauf, '92; Schubert et al., '93). They specify developmental cell fates along the anterior–posterior axis in bilaterian animals.

Grant sponsors: Max-Planck-Gesellschaft, Deutsche Forschungsgemeinschaft and Natural Science Foundation of China.
*Correspondence to: A. Meyer, Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, Universitätsstr. 10, 78457 Konstanz, Germany.
E-mail: axel.meyer@uni-konstanz.de
Received 19 June 2006; Revised 8 October 2006; Accepted 11 October 2006
Published online 29 January 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jez.b.21144.

The increased complexity of body plans that has accompanied the evolution of higher vertebrates is a phenomenon of particular interest and importance (Martinez and Amemiya, 2002) and a particularly intriguing problem is the understanding of the role of Hox cluster duplications in the evolution of vertebrates. In this regard, questions such as the characterization of evolutionary patterns after the FSGD including Hox gene retention/deletion, and evolution at the nucleotide level is of particular interest (Holland et al., '94; Malaga-Trillo and Meyer, 2001; Wagner et al., 2003; Prohaska and Stadler, 2004; Hoegg and Meyer, 2005; Crow et al., 2006; Kurosawa et al., 2006). Available DNA sequences from teleosts have shown a variety of gene retention and deletion patterns. These might possibly be relevant to the adaptive evolution during the initial teleost radiation, yet the role of the cluster duplications in vertebrate evolution is still poorly understood, especially the evolution immediately after the duplication (Wagner et al., 2003, 2005).

Genome duplications have been an exceedingly rare phenomenon in animals, yet they occurred somewhat more frequently in some vertebrate lineages, such as fishes, than in others (Ferris and Whitt, '77; Allendorf, '78; Wolfe, 2001). The short-term impact of genome duplications is still not well understood. How is dosage balance achieved when the genome suddenly contains additional copies of a gene? Such questions have been explored in plants (Soltis and Soltis, 2000; Wendel, 2000; Chen et al., 2004; Comai, 2005). What are the mechanisms leading to retention or loss of a duplicated gene? A partial answer to the latter questions is provided by the DDC model (Force et al., '99; Lynch et al., 2001). It is not clear, however, whether the situation in plants carries over to vertebrates (Furlong and Holland, 2004; LeComber and Smith, 2004). For instance, it remains unclear how important developmental regulators, such as Hox genes, have evolved initially after gene or genome duplications. More precisely, it remains unknown how quickly and how much silencing, neofunctionalization, or subfunctionalization takes place in the aftermath of genome duplications. Comparative studies have shown that gene loss from Hox gene clusters is an ongoing process, i.e., the resolution of the post-duplication redundancy is not immediate (Prohaska and Stadler, 2004; Hoegg and Meyer, 2005).

Polyploidy occurs in several unrelated groups of fish such as salmonids, sturgeons, and cyprinids and might have played an important role in regulatory evolution (LeComber and Smith, 2004). The importance of genome duplications in shaping the evolution of genomes can best be examined by comparative genomic analyses of sequences from several closely related organisms that vary in the number of genome duplications their lineages have experienced. In this regard, closely related diploid and tetraploid organisms provide the cleanest test situation for the investigation of ''post-ploidy'' events. Those pairwise species comparisons can be used to test the effects of those drastic genomic events, in terms of dosage balance or fast stabilization of duplicated genomes via retention/exclusion of redundant genomic DNA regions.

More than 40 species in the family Cyprinidae (mainly in the three subfamilies Cyprininae, Schizothoracinae, and Barbinae, but rare in the other subfamilies), are known to have undergone repeated genome duplication events (Chen et al., '84; Yu et al., '89; Chen, '98; Li et al., personal communication, Luo et al., unpublished data). Various types of polyploidy have been observed: tetraploidy ($N = 4$; goldfish and common carp; $2n = 100$), hexaploidy ($N = 6$; *Schizothorax prenanti* Tchang; $2n = 148$), octaploidy ($N = 8$; *Carassius auratus gebelio* and *C.a. langsdorfi*; $2n = 200\pm$) and, although rarely, triploidy ($N = 3$; *Phoxinus eosneogaeus*), as well (Dawley and Goddard, '88; Yu et al., '89; Murakami et al., 2001; He et al., personal communication).

Goldfish belong to the subfamily Cyprininae (Howes, 1991). The ancestor of all species in this subfamily is thought to have been tetraploid (Yu et al., '89). The most recent genome duplication event(s) in this subfamily of fish are believed to have occurred within the last 20 million years (Risinger and Larhammar, '93; Yang and Gui, 2004). In addition, recent and recurrent hexaploids were discovered in different lineages as well (Luo et al., unpublished data), which makes the goldfish and its relatives a particularly interesting group in which to investigate the short-term evolutionary effects of genome duplications since their duplicated genomes are still within the half life of duplicates as suggested by Lynch and Conery (2000). This group thus appears to be an excellent model with which to test hypotheses relevant to the vertebrate genome duplication and diversification.

In particular, the fate of extra copies of crucial developmental regulators such as the Hox genes can be expected to shed some light on the mechanisms that act in duplicated genomes after

a genome duplication. As a first step towards addressing these issues, a PCR survey of Hox genes in the goldfish *C.a. auratus*, a putative young allo-tetraploid species, was performed with the goal of determining whether a genomic screening of Hox clusters in a goldfish BAC genomic library (Luo et al., 2006) is likely to be fruitful. The PCR data already provided some interesting insights into the post-duplication evolution of goldfish Hox gene clusters.

MATERIALS AND METHODS

One male orange specimen of *C.a. auratus* was obtained commercially in Konstanz, Germany and used for genomic DNA extraction. Hox genes were amplified from its genomic DNA. The detailed extraction procedure is described in Luo et al. (2006).

Degenerate primers for the amplification of Hox genes included the posterior Hox forward primers for paralogous groups 9–13 [5′CGA AAG AAG (C/A) G(N/C)GT(N/C) CC(N/C)T A(T/C)AC, anterior Hox forward primer for paralogous groups 1–9 [GAA TTC CAC TTC AAC(C/A)(G/A)(C/G) TAC CT], and the universal reverse primer [CAT CCT GCG GTT TTG GAA CCA NAT], as described by Amores et al. (2004).

PCR reactions were conducted in a volume of 10 μl PCR cocktail that included 1 × buffer with 0.15 mmol $MgCl_2$ (Sigma), 0.25 mM dNTPs (Fermantas), 0.5 U Taq DNA polymerase (Sigma), and 12–15 ng genomic DNA. Following a 2 min denaturing period at 94°C, the following PCR conditions were used for the experiments both with HOX1–9 primer+universal reverse primer and HOX9–13 primer+universal reverse primer: 35 cycles at 94°C for 30 sec, 50°C for 30 sec, and 72°C for 60 sec, followed by a final extension at 72°C for 5 min.

Resulting PCR products were purified with purification columns (Qiagen). Then the purification products were ligated and cloned with the TA cloning kit (Invitrogen). Each clone was amplified again with the universal M13 forward and reverse primers and sequenced with an ABI 3100 automatic sequencer, by using an ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit (with AmpliTaq DNA polymerase FS, Applied Biosystems).

Hox sequences have been deposited in Genbank with accession numbers DQ630465–DQ630513 for all unique sequences. Some clones shared the same sequence, and the clone number information was included in the Genbank entry.

The sequences of 73 PCR clones and their reverse complements were aligned using ClustalW (Thompson et al., '94) to identify identical clones. Blastn searches in Genbank were used to reconfirm that the sampled sequences were not contaminations. The DNA sequences were then translated in all six possible reading frames and compared to known homeodomain-containing proteins to obtain the putative Hox protein sequence fragments. Several PCR products had one or two deletions, a single clone (H01 56 15) had two inserted nucleotides relative to the standard homeobox sequence. Based on peptide sequence identity with known Hox genes, DNA sequences could be assigned unambiguously to paralog groups (PG) 1, 3, 9, 11, 13, as well as to a more poorly resolved ''middle group'' gene (paralog groups 4–7). The automated phylogenetic key for classifying homeoboxes described by Sarkar et al. (2002) unfortunately is not applicable since it requires homeobox fragments that are longer than 27 amino acids.

11 PCR clones did not correspond to Hox genes, and were tentatively identified as two copies of the HB9-type homeobox gene hlxb9la. The remaining 62 Hox homeobox sequences were analyzed further. In addition, we re-analyzed the 11 previously published Hox homeobox fragments by Levine and Schechter ('93) (see Table 1).

Due to the relatively close phylogenetic relationship between *C.a. auratus* and *Danio rerio*, most but not all PCR products could be unambiguously identified as orthologs of one of the 49 zebrafish Hox genes. To this end, we measured the Hamming distance Δ of the goldfish Hox sequences (as well as the sequences from Levine and Schechter ('93)) to each of the zebrafish sequences. For comparison, we determined for each zebrafish Hox fragment the Hamming distance D to the closest other Hox fragment. If a *C.a. auratus* sequence x was significantly closer to a unique *D. rerio* x' sequence than this *D. rerio* sequence x' was to any other *D. rerio* sequence y', we could conclude that x was the true ortholog of x'. More formally, we assumed that goldfish sequence x was homologous to zebrafish sequence x' if $D = d(x, x') << \min_{y', x'} d(x', y') = \Delta$. For the computation of the Hamming distance we treat gaps as fifth character. Since gaps only occur in the pseudogenes, this simple choice does not affect the assignment of genes to paralog groups. Through this method we could identify

TABLE 1. Summary of C.a. auratus PCR fragments

PG	pg	*D*	Δ	Previous work[1]	Copies	Original clone IDs
1	A1a	9	22	L09698	—	
	B1b	7–8	18		2	C04′27′06, F01′6′11
2	A2b	7	18	L09691	—	
	A2a	*	*	L09690	—	Pseudogene in *D. rerio*
3	*B3a?	10–13	15	L09697	6	H′0540′15, B04′26′04, E05′37′09, F02′14′12, F03′22′11, G04′31′14
	*A3a?	14–15	19		2	C05′35′05, G01′7′13
4	C4a	7	20		2	A03′17′01, C01′3′05
5	A5a	4–8	18	L09688	1	B03′18′03
	A5a?				1	C03′19′05 artifact?
	B5a	2	13	L09686	—	
	B5b	4–5	13	L09685/9	—	
	C5a	4	13		2	B06′42′04, G03′23′13
6	B6b	3	15	L09687	—	
	C6a	10	19		2	A05′33′01, H04′32′16
7	B7a	10–12	21		2	D05′36′07, H03′24′15
9	A9b	4	19		1	D04′76′08
10	A10b	6–9	17	L09694	6	A04′73′02, H02′64′16, A03′65′01, F05′38′12, F04′78′12, D03′20′07
		11	22		4	[long] A04′73′02, H02′64′16′, A03′65′01, F04′78′12
	*B10a	7	16		4	A04′25′02, A06′41′02, B01′2′03, A02′9′02
	C10a	0	15	L09693	—	
	D10a	7	15		1	B05′34′03
13	A13a	8	16		1	G04′79′14
		11	23		1	[long] -″-
	A13b	3	16		6	H06′48′16, E01′53′09, F03′70′11, G01′55′13 C04′75′06, E03′69′09
		5	22		6	[long] -″-
	B13a	3	22		11	A02′57′02, D01′52′07, G03′71′13, A01′49′01 D03′68′07, F02′62′12, E02′61′10, G02′63′14 H03′72′15, E04′77′10, F01′54′11
		5–6	29		11	[long] -″-
	C13a	4	16		10	C06′43′06, D06′44′08, F06′46′12, B01′50′03 B02′58′04, B04′74′04, C02′59′06, D02′60′08 C01′51′05, H04′80′16
		7	25		10	[long] -″-
	*D13a	16	30		1	[long] two extra A inserted H01′56′15
	?				1	G06′47′14

PG, number of Hox paraglogs; pg, assigned Hox paraglog; *, putative pseudogene in *C.a. auratus*; ?, uncertain assignment of gene identity from Hamming distances; [long], 114nt sequences instead of 81nt sequences for the posterior PGs; *D*, Hamming distance between goldfish PCR fragment and nearest zebrafish gene; Δ, Hamming distance between most closely related zebrafish genes.
[1]Levine and Schechter ('93).

the orthology relationships of most of the PCR fragments (Table 1).

The identities of the PCR fragments were also investigated by reconstructing phylogenetic trees using different methods following the procedure outlined by Prohaska and Stadler (2004). Neighbor-joining (Saitou and Nei, '87), parsimony, and maximum likelihood analysis for each PG was performed using the Phylip package (Felsenstein, '89), using other teleost Hox homeobox fragments for comparison (*D. rerio*, *Tetraodon nigroviridis*, *Takifugu rubripes*, *Oryzias latipes*, and *Fundulus heteroclitus*). In most cases goldfish sequences are grouped with a unique zebrafish sequence in these trees. Bootstrap support for this grouping is at least 70% in all cases not marked by ''?'' in Table 1. Furthermore, all assignments derived from the trees are consistent with those listed in Table 1. Phylogenetic trees as well as neighbor nets (Bryant and Moulton, 2004) (computed using SplitsTree of Huson ('98), based on the Jukes–Cantor distance) are compiled in the electronic supplement (URL: http://www.bioinf.uni-leipzig.de/Publications/SUPPLEMENTS/06-007/). We use the neighbor-net approach in addition to classical phylogenetic methods as it provides a convenient graphical representation of noise and ambiguities in the data. Quartet Mapping (Nieselt-Struwe and

von Haeseler, 2001), as implemented in the quartm 0.2 program (Stadler et al., 2004), was used to provide further corroborating evidence (data not shown), again with consistent results. In this method, we treated gaps as missing data: for each quadruple we ignored all those columns in the alignment in which at least one gap was contained in one of the four sequences under consideration.

RESULTS

The combined dataset of our PCR products and the sequences reported in Levine and Schechter ('93) yield 26 distinct Hox gene fragments in the goldfish *C.a. auratus*. The orthology relationships of all but two genes could be identified with confidence (Table 1). We observe sequence variations from different clones that were apparently amplified from a single orthologous Hox gene (also Table 1). Due to the short sequence length it is not possible to determine whether this variation is due to allelic variants, divergence of duplicate genes (i.e., paralogous genes) following the recent tetraploidization, or whether it might be due to potential sequencing artifacts. Sequence variation is concentrated, however, in a small number of sequence groups, suggesting relaxed selection in at least one copy of the duplicated genes.

Two groups of Hox3 sequences seem to be recent pseudogenes as judged by the existence of indels in the homeobox sequences. An alignment of the putative HoxB3a sequences is shown in Figure 1. On the other hand, expression of L09697 (G11-4), which is indistinguishable from the HoxB3a PCR products, has been reported in the goldfish brain (Levine and Schechter, '93). The discrepancy is that only one copy of HoxB3a in the tetraploid goldfish might have turned into a pseudogene.

Similarly, various deletions suggest that HoxB10a is a pseudogene in goldfish and for the HoxA3a gene we found only two sequences with a deletion at the same position (not shown). A single clone similar to zebrafish HoxD13a has two inserted A residues as well as a conspicuous run of T nucleotides, strongly suggesting that it is either a PCR artifact or a real pseudogene.

Since we consistently find indels in independent amplifications belonging to the same Hox genes, we conclude that PCR or cloning-induced errors are an unlikely explanation for the observed ''corrupt'' homeobox sequences. Rather, at least both Hox3 paralogs as well as HoxB10a are most likely true pseudogenes.

It has been reported previously that homeobox fragment L09690 (G5-1) corresponds to a rare 1.4 kb transcript that is expressed in the goldfish brain (Levine and Schechter, '93). Comparison with other Hox2 sequences (Fig. 2) shows that this sequence is homologous to the HoxA2a pseudogene of zebrafish. This observation implies that HoxA2a was lost in zebrafish after the split of the *Danio* and *Carassius* lineages.

We are thus left with two single PCR clones for which no detailed assignment can be made. One of them belongs to PG13. By comparison with the Hox gene complement of the zebrafish, it is tempting to speculate that this is the ''missing'' *HoxC13b* gene. The second sequence belongs to the middle group and differs by 13 nucleotides from *HoxA5a*.

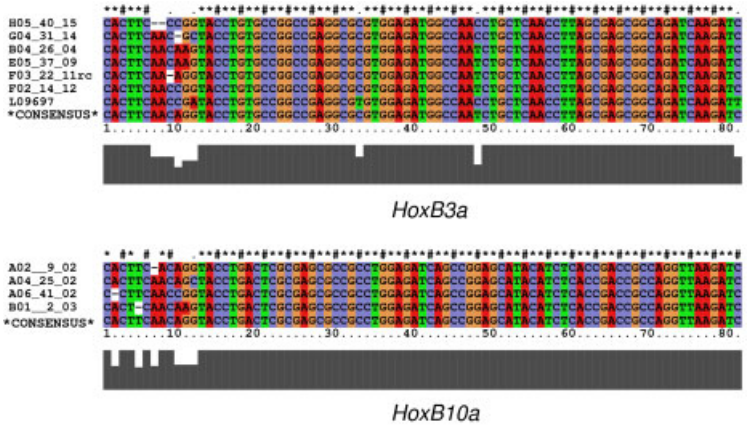

Fig. 1. Frequent indels suggest that HoxB3a and HoxB10a have recently turned into pseudogenes. Names refer to original clone numbers in Table 1.

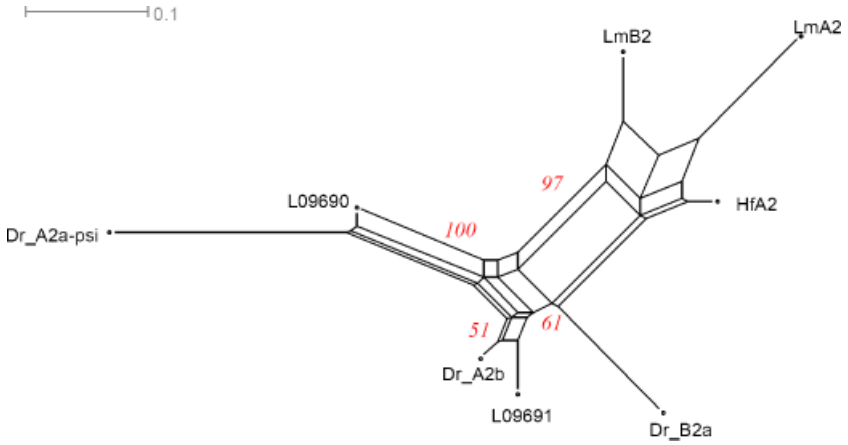

Fig. 2. Neighbor-net analysis of Hox2 genes strongly suggests that L09690 (Levine and Schechter, '93) is orthologous to the zebrafish HoxA2a pseudogene. Dr-*Danio rerio*, Hf-*Heterodontus francisci*, Lm-*Latineria menadonesis*.

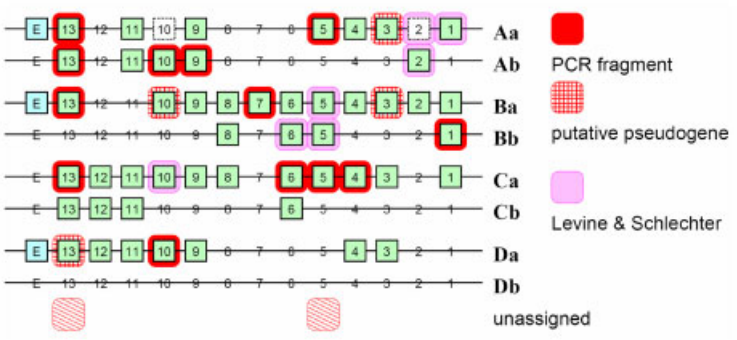

Fig. 3. Goldfish Hox genes in comparison to zebrafish. Our data are largely consistent with the expectation that goldfish and zebrafish should have a similar organization of Hox clusters. Differences are essentially restricted to recent pseudo-gene formations of HoxA2a in *Danio* but not in *Carassius*, while HoxB10a, HoxB3a, and likely also HoxA3a have turned into pseudogenes in the goldfish. Green boxes represent Hox genes of *Danio rerio*, and "E" refers to Evx (even-skipped related gene).

DISCUSSION

The results of the PCR survey data and the published sequences from Levine and Schechter ('93) are summarized in Figure 3 in comparison to the Hox cluster complement of the zebrafish. *D. rerio* has 49 Hox genes in seven tightly linked clusters, all of which are expressed in the adult fish (Corredor-Ad'amez et al., 2005). In addition, there are at least two pseudogenes, HoxA2aψ and HoxA10aψ in the zebrafish Hox clusters. Not surprisingly, the homeobox fragments reported here indicate that the Hox cluster organization of goldfish is similar, but not identical to that of zebrafish.

Surprisingly, the pseudogenes found in the Hox clusters of zebrafish and goldfish are different. While at least one of the zebrafish pseudogenes, the pseudogene of HoxA2a, is expressed in goldfish, the PCR results strongly suggest that several intact zebrafish genes: HoxB10a, Hox3a, and possibly also HoxA3a might have evolved into pseudogenes in the goldfish lineage. In the Hox10 paralog group, zebrafish has lost HoxA10a completely, while in the goldfish HoxB10a has recently turned into a pseudogene. Even taking into account that our PCR survey covers only 19 (or 25, if we include the data from Levine and Schechter ('93)) of the expected roughly 50 Hox genes, we already detect strong indications that pseudogene formation is much more prevalent in goldfish (3–5 of 25 PCR fragments) than in zebrafish (2 of 51 genes).

Dosage effects might plausibly cause selection pressure to lose redundant Hox genes through the formation of pseudogenes. In this context, it is intriguing that our examples of pseudogenes are recognizable mostly because they contain indels in

the homeobox, which would almost certainly render the gene non-functional even if it was still expressed.

Pseudogenes are formed by random mutations that create either stop codons which prematurely terminate the full-length functional expression product, or by insertion/deletions causing a shift of the reading frame, thereby rendering the translated protein non-functional. Although the substitution pattern in pseudogenes is not completely random, pseudogenes are under nearly selective neutrality (Gojobori and Li, '82; Li et al., '84; Li, '97). In theory, the pseudogene formation happens as likely via both stop codon creation as indels. Yet in our study of the goldfish Hox pseudogenes, indels seem to be dominant. This biased pattern might arise from the selective pressure of removing the gene product soon after the duplication event, a potential effect that will require further investigation.

Moghadam et al. (2005a,b) have indicated there were at least 14 Hox gene clusters in trout and salmon (family Salmonidae). This group of fish was thought to have evolved from an ancestor in which an autotetraploidization event occurred 25–100 Mya (Ohno, '70; Allendorf and Thorgaard, '84). If goldfish has retained most of its duplicated genome following the recent tetraploidization, we would expect 14–16 Hox gene clusters, a similar number to the salmon fishes (Risinger and Larhammar, '93; Amores et al., 2004; Yang and Gui, 2004). Also, due to the recent tetraploidization event we expect that only a few mutations accumulated between the homeobox sequences of the paralogs. As a consequence, we cannot distinguish with certainty recent paralogs from allelic variants, and hence cannot determine a minimum number of retained clusters at present.

Luo et al. (2006) identified two copies of the recombinase-activating gene 1 (RAG1) from a goldfish genomic BAC library, while in zebrafish only a single-copy gene exists. Divergence analysis of RAG1 dated the gene duplication 14.2–14.5 Mya, in agreement with previous dating of the goldfish-specific gene duplication (Risinger and Larhammar, '93; Yang and Gui, 2004), it is likely that the duplication of RAG1 was caused by the genome duplication event in the goldfish lineage. This suggests that the initially functional genes were turned into pseudogenes within the last 15 Mya. This age estimation is consistent with the relative small branch lengths that still allow a confident classification of the goldfish Hox genes.

In summary, our PCR survey demonstrates that there is a high rate of pseudogene formation dominated by indels in the pseudogenes of goldfish Hox genes. This is consistent with the expectation that there are more Hox clusters in the goldfish genome which have lost functional Hox genes, thereby reducing redundancy following the recent tetraploidization event. Due to the small evolutionary distance, however, extensive sequencing of BAC clones of one individual containing goldfish Hox clusters will be necessary to obtain a more complete understanding of the processes involved in this secondary gene loss and immediate evolution after duplication.

ACKNOWLEDGMENTS

We thank Kathrin Hoffmann for technical assistance and Simone Hoegg, Nicol Siegel, Ingo Braasch, Walter Salzburger, Elke Hespeler, and Sonja J. Prohaska for valuable discussions. PFS thanks the CAS/MPG Partner Institute at the SIBS in Shanghai for its hospitality in March 2006, where part of the analysis was performed. This work has been funded, in part, by the Deutsche Forschungsgemeinschaft, through grants to A.M. and the DFG Bioinformatics Initiative (P.F.S.), by the Max-Planck-Society (A.M., J.L.), and the Natural Science Foundation of China (S.H., J.L.).

LITERATURE CITED

Allendorf FW. 1978. Protein polymorphism and the rate of loss of duplicate gene expression. Nature 272:76–78.

Allendorf FW, Thorgaard GH. 1984. Tetraploidy and the evolution of salmonid fish. In: Turner JB, editor, Evolutionary genetics of fish. New York: Plenum Press. p 1–53.

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-fin fish. Genome Res 14:1–10.

Bryant D, Moulton V. 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol Biol Evol 21:255–265.

Chen XL, Yue PQ, Lin RR. 1984. Major groups within the family Cyrinidae and their Phylogenetic relationships. Acta Zootaxon Sin 4:424–440.

Chen YY. 1998. Preface. In: Chen Yiyu et al., editors, Fauna Sinica, Osteichthyes, Cypriniformes II. Beijing: Science Press. p 1–18.

Chen ZJ, Wang J, Tian L, Lee HS, Wang JJ, Chen M, Lee JJ, Josefsson C, Madlung A, Watson B, Pires JC, Lippman Z, Vaughn MW, Colot V, Birchler JA, Doerge RW, Martienssen R, Comai L, Osborn T. 2004. The development of an

Arabidopsis model system for genome-wide analysis of polyploidy effects. Biol J Linn Soc 82:689–700.
Comai L. 2005. The advantages and disadvantages of being polyploid. Nat Rev Genet 6:836–846.
Corredor-Ad'amez M, Welten MC, Spaink HP, Jeffery JE, Schoon RT, de Bakker MA, Bagowski CP, Meijer AH, Verbeek FJ, Richardson MK. 2005. Genomic annotation and transcriptome analysis of the zebrafish (*Danio rerio*) hox complex with description of a novel member, HoxB13a. Evol Dev 7:362–375.
Crow KD, Stadler PF, Lynch VJ, Amemiya CT, Wagner GP. 2006. The fish specific Hox cluster duplication is coincident with the origin of teleosts. Mol Biol Evol 23:121–136.
Dawley RM, Goddard KA. 1988. Diploid–triploid mosaics among unisexual hybrids of the minnows *Phoxinus eos* and *Phoxinus neogaeus*. Evolution 42:649–659.
Dehal P, Boore JL. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol 3: e314.
Donoghue PC, Purnell MA. 2005. Genome duplication, extinction and vertebrate evolution. Trends Ecol Evol 20: 312–319.
Felsenstein J. 1989. Phylip–phylogeny inference package (version 3.2). Cladistics 5:164–166.
Ferris SD, Whitt GS. 1977. Loss of duplicate gene-expression after poly-ploidization. Nature 265:258–260.
Force A, Lynch M, Pickett FB, Amores A, Yan Yl, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545.
Furlong RF, Holland PWH. 2004. Polyploidy in vertebrate ancestry: Ohno and beyond. Biol J Linn Soc 82:425–430.
Gojobori T, Li WH. 1982. Patterns of nucleotide substitution in pseudogenes and functional genes. J Mol Evol 18: 360–369.
Gregory TR, Hebert PDN. 1999. The modulation of DNA content: proximate causes and ultimate consequences. Genome Res 9:317–324.
Hoegg S, Meyer A. 2005. Hox clusters as models for vertebrate genome evolution. Trends Genet 21:421–424.
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.
Holland PWH, Garcia-Ferńandez J, Williams NA, Sidow A 1994. Gene duplication and the origins of vertebrate development. Development (Suppl.):125–133.
Howes GJ. 1991. Systematics and biogeography: an overview. In: Winfield IJ, Nelson JS, editors. Cyprinid fishes, systematics, biology and exploitation. London: Chapman and Hall. p 1–33.
Huson DH. 1998. Splitstree: analyzing and visualizing evolutionary data. Bioinformatics 14:68–73.
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.
LeComber S, Smith C. 2004. Polyploidy in fishes. Biol J Linn Soc 82:432–442.
Levine EM, Schechter N. 1993. Homeobox genes expressed in the retina and brain of adult goldfish. Proc Natl Acad Sci USA 90:2729–2733.
Li WH. 1997. Gene structure, genetic codes, and mutation. In: Li WH, editor, Molecular evolution, Sunderland MA: Sinauer Associates. p 31–33.
Li WH, Wu C, Luo C. 1984. Nonrandomness of point mutations as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. J Mol Evol 21:58–71.
Luo J, Lang M, Salzburger W, Siegel N, Stoelting K, Meyer A. 2006. A BAC li-brary for the goldfish *Carassius auratus auratus* (Cyprinidae, Cypriniformes). J Exp Zool B, in press.
Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. Science 290: 1151–1155.
Lynch M, O'Hely M, Walsh B, Force A. 2001. The probability of preservation of a newly arisen gene duplicate. Genetics 159:1789–1804.
Malaga-Trillo E, Meyer A. 2001. Genome duplications and accelerated evolution of Hox genes and cluster architecture in teleost fishes. Am Zool 41:676–686.
Martinez P, Amemiya CT. 2002. Genomics of the HOX gene cluster. Comp Biochem Physiol B: Biochem Mol Biol 133: 571–580.
McGinnis W, Krumlauf R. 1992. Homeobox genes and axial patterning. Cell 68:283–302.
Meyer A, Van de Peer Y. 2005. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). Bioessays 27: 937–945.
Moghadam HK, Ferguson MM, Danzmann RG. 2005a. 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.
Moghadam HK, Ferguson MM, Danzmann RG. 2005b. Evidence for Hox gene duplication in rainbow trout (*Oncorhynchus mykiss*): a tetraploid model species. J Mol Evol 61:804–818.
Murakami M, Matsuba C, Fujitani H. 2001. The maternal origins of the triploid ginbuna (*Carassius auratus langsdorfi*): phylogenetic relationships within the *C. auratus* taxa by partial mitochondrial d-loop sequencing. Genes Genet Syst 76:25–32.
Nelson J. 1994. Fishes of the world. New York: Wiley.
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.
Ohno S, Atkin NB. 1966. Comparative DNA values and chromosome complements of eight species of fishes. Chromosoma 18:455–466.
Ohno S, Muramoto J, Christian L, Atkin NB. 1967. Diploid–-tetraploid relationship among old-world members of the fish family cyprinidae. Chromosoma 23:1–9.
Prohaska SJ, Stadler PF. 2004. The duplication of the hox gene clusters in teleost fishes. Th Biosci 123:89–110.
Risinger C, Larhammar D. 1993. Multiple loci for synapse protein SNAP-25 in the tetraploid goldfish. Proc Natl Acad Sci USA 90:10598–10602.
Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425.
Sarkar IN, Thornton JW, Planet PJ, Figurski DH, Schierwater B, DeSalle R. 2002. An automated phylogenetic key

for classifying homeoboxes. Mol Phylogenet Evol 24: 388–399.

Schubert FR, Nieselt-Struwe K, Gruss P. 1993. The antennapedia-type home-obox genes have evolved from three precursors separated early in metazoan evolution. Proc Natl Acad Sci USA 90:143–147.

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

Soltis P, Soltis DE. 2000. The role of genetic and genomic attributes in the success of polyploids. Proc Natl Acad Sci USA 97:7051–7057.

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 Phylogenet Evol 32:686–692.

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, Braasch I, Meyer A. 2001. Comparative genomics provides evidence for an ancient genome duplication event in fish. Philos Trans R Soc Lond Ser B 356:1661–1679.

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. Nuclei Acids Res 22:4673–4680.

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

Wagner A. 2001. Birth and death of duplicated genes in completely sequenced eukaryotes. Trends Genet 17: 237–239.

Wagner GP, Amemiya C, Ruddle F. 2003. Hox cluster duplications and the opportunity for evolutionary novelties. Proc Natl Acad Sci USA 100:14603–14606.

Wagner GP, Takahashi K, Lynch V, Prohaska SJ, Fried C, Stadler PF, Amemiya C. 2005. Molecular evolution of duplicated ray finned fish HoxA clusters: increased synonymous substitution rate and asymmetrical co-divergence of coding and non-coding sequences. J Mol Evol 60:665–676.

Wendel JF. 2000. Genome evolution in polyploids. Plant Mol Biol 42:225–249.

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

Wolfe K. 2001. Yesterday's polyploidy and the mystery of diploidization. Nat Rev Genet 2:333–341.

Yan YL, Willoughby J, Liu D, Crump JG, Wilson C, Miller CT, Singer A, Kimmel C, Westerfield M, Postlethwait JH. 2005. A pair of Sox: distinct and overlappig functions of zebrafish sox9 co-orthologs in craniofacial and pectoral fin development. Development 132:1069–1083.

Yang L, Gui JF. 2004. Positive selection on multiple antique allelic lineages of transferrin in the polyploid *Carassius auratus*. Mol Biol Evol 21:1264–1277.

Yu XJ, Zhou T, Li YC, Li K, Zhou M. 1989. Chromosomes of Chinese fresh-water fishes. Beijing: Science Press.