Molecular Evolution of Duplicated Ray Finned Fish HoxA Clusters:
Increased synonymous substitution rate and asymmetrical co-divergence of coding and non-coding sequences.

Günter P. Wagner*+, Kazuhiko Takahashi*, Vincent Lynch*, Sonja J. Prohaska#, Claudia Fried#, Peter F. Stadler#, and Chris Amemiya%

*)Department of Ecology and Evolutionary Biology,
Yale University, New Haven, CT 06520-8106, USA
#) Bioinformatik, Institut für Informatik,
Universität Leipzig, D-04103 Leipzig, Germany
%) Benaroya Research Institute at Virginia Mason,
Seattle, WA 96101, USA

+) Corresponding author:
Gunter P. Wagner
Department of Ecology and Evolutionary Biology
Yale University
New Haven, CT 06520-8106
Tel: 203-432-9998
Fax: 203-432-3870
Email: gunter.wagner@yale.edu

Abstract:

In this study the molecular evolution of duplicated HoxA genes in zebrafish and fugu has been investigated. All eighteen duplicated HoxA genes studied have a higher non-synonymous substitution rate than the corresponding genes in either bichir or paddlefish, where these genes are not duplicated. Surprisingly, though, the higher rate of evolution is not due solely to a higher non-synonymous to synonymous rate ratio but to an increase in both the non-synonymous as well as the synonymous substitution rate. The synonymous rate increase can neither be explained by changes in nucleotide composition nor codon bias. Thus, we suggest that the HoxA genes may experience an increased mutation rate following cluster duplication. In the non-Hox nuclear gene *RAG1* only an increase in non-synonymous substitutions could be detected, suggesting that the increased mutation rate is specific to duplicated Hox clusters and might be related to the structural instability of Hox clusters following duplication. The divergence among paralog genes tends to be asymmetric with one paralog diverging faster than the other. In fugu, all b-paralogs diverge faster than a-paralogs, while in zebrafish *Hoxa-13a* diverges faster. This asymmetry corresponds to the asymmetry in the divergence rate of conserved non-coding sequences, i.e. putative cis-regulatory elements. In fugu, the putative cis-regulatory elements of the b-paralog HoxA cluster diverge faster, while in zebrafish those of the a-paralog cluster diverge faster. Hence, we found evidence for a concerted asymmetric divergence of coding sequences on the same cluster as well as between coding and non-coding sequences. These results suggest that the 5' HoxA genes in the same cluster belong to a co-evolutionary unit in which genes have a tendency to diverge together.

Introduction

Hox cluster duplications play a prominent but poorly understood role in vertebrate evolution (Wagner et al. 2003). Hox genes code for homeodomain containing transcription factors and are homologous to the homeotic genes in the Drosophila Antp and Ubx gene clusters. The Hox gene family is thought to have arisen by tandem duplications, leading to a cluster of three to four homeobox containing genes, which was then duplicated, giving rise to the ancestral Hox and para-Hox gene clusters (Holland and Garcia-Fernandez 1996; Kappen et al. 1989). Additional tandem duplications increased the number of Hox genes in the cluster to up to 14 in chordates (Ferrier et al. 2000; Powers and Amemiya 2004). All well characterized invertebrate taxa have a single cluster with the exception of the nematode, *Caenorhabditis elegans*, and the urochordate *Ciona intestinalis* which have lost the integrity of the Hox cluster all together (Dehal and et 2002; Ruvkun and Hobert 1998). In contrast, multiple copies of Hox clusters have been found in all vertebrate lineages, varying between at least three in lamprey to seven or eight in teleost species (Holland et al. 1994; Martinez and Amemiya 2002; Meyer and Malaga-Trillo 1999; Prohaska and Stadler 2004; Ruddle et al. 1994b; Taylor et al. 2003). From these data at least three duplication events can be inferred, although the number of duplications might be as high as nine (Fried et al. 2003). The reason for the different pattern of Hox cluster evolution in vertebrate compared to invertebrate lineages is not clear. One possibility is that the duplicated genes are necessary for vertebrate body plan elaboration (Holland et al. 1994; Malaga-Trillo and Meyer 2001; Ruddle et al. 1994a), or to overcome structural constraints specific to vertebrate Hox clusters (Fried et al. 2004; Wagner et al. 2003). In this paper, we examine the evolutionary fate of duplicated HoxA genes in the teleosts in order to gain insight into the role Hox cluster duplication plays in vertebrate evolution.

A variety of teleost species are known to have more than four Hox clusters (Meyer and Schartl 1999; Prince and Pickett 2002; Prohaska and Stadler 2004), including the zebrafish(Amores et al. 1998), Fugu and Spheroides (Amores et al. 2004), and medaka (Naruse et al. 2000; Naruse et al. 2004). Preliminary data also exists for the killifish (Misof and Wagner 1996), stripped bass (Pavell and Stellwag 1994; Snell et al. 1999), tilapia (Santini et al. 2003), and the rice field eel (Ji et al. 2002). It is not entirely clear when this duplication has happened, but data suggests that it was some time after the most recent common ancestor of paddlefish and other ray fined fishes and before the most recent common ancestor of Euteleosts (Metscher et al. 2004).

Three sets of taxa are considered in this paper, one for which Hox gene sequences are available and one where RAG1 sequence information is known and the non-coding sequences of zebrafish, fugu, tilapia, bichir and shark HoxA clusters. The phylogenetic hypotheses for these two sets of species are given in Figures 1a and 1b.

Materials and Methods

Data:
Full Hox cluster sequences were used in this study from the shark, *Heterodontus francisci*: HfHoxA = AF479755; the bichir, *Polypterus senegalus* : PsHoxA = AC132195 and AC12632; the two zebrafish paralogs, *Danio rerio*: DrHoxAa =

AC107365 r.c.; and DrHoxAb = AC107364; the fugu, *Takifugu rubripes:* TrHoxAa = Fugu v.3.0 scaffold 47 positions 103001-223000 r.c., contains FRU92573; and TrHoxAb = Fugu v.2.0 scaffold 1874, and tilapia, *Oreochromis niloticus*: OnHoxA: AF533976.

The *Hoxa-13* and *Hoxa-11* sequences of paddlefish were PCR amplified, cloned and sequenced and described elsewhere (Metscher et al. 2004). The sequences are published in Genbank (Hoxa-13: # AY661749, protein ID AAT75331) (Hoxa-11: genomic sequence # AY661748, protein ID AAT75330).

RAG1 sequences of the following species where retrieved from Genbank: zebrafish, *Danio rerio* (NM131389), fugu, *Takifugu rubripes* (AF108420), winter flounder, *Pseudopleuronectus americanus* (AF369067), Atlantic cod, *Gadus morhua* (AF369064); Osteoglossomorpha: bronze beatherback, *Notopterus notopterus* (AF369063), *Ganthonemus sp*. (IMCB-2001) (AF369062), freshwater butterfly fish, *Pantodon buchholzi* (AF369061), *Osteoglossum sp*. (AF369060); bowfin, *Amia calva* (AF369059), Mississippi paddlefish, *Polyodon spathula* (AF369057), bichir, *Polypterus sp*. (IMCB-2001) (AF369055), *Latimeria menadoensis* (AF369069).

Analysis of coding sequence evolution: both the nucleotide sequence and predicted amino acid sequence were aligned with ClustalW. The amino acid sequence was then inspected for regions of ambiguous alignment. These regions and gaps were eliminated from the alignment. Finally, the nucleotide sequence corrected to conform to the edited amino acid sequence. Sequence alignments were analyzed with three programs as appropriate and indicated in the results section: RRTree 1.1.11 (Robinson-Rechavi and Huchon 2000), HYPHY99.beta for MacOS (Muse and Gaut 1994), and PAML (Yang 1997).

Relative Rate Test for Conserved Non-Coding Sequences: conserved non-coding sequences were detected using the `tracker` program (Prohaska et al. 2004). Very briefly, this approach is based on `BLAST` (Altschul et al. 1990) for the initial search of all pairs of input sequences restricted to homologous intergenic regions. The resulting list of pairwise sequence alignments is then assembled into groups of partially overlapping regions that are subsequently passed through several filtering steps and finally aligned using the segment based multiple alignment tool `DIALIGN2` (Morgenstern 1999). The final output of the program is the list of these aligned "footprint cliques" (see supplemental material at http://www.tbi.univie.ac.at/Publications/SUPPLEMENTS/04-014). The alignments of all footprint cliques are concatenated and padded with gap characters where a footprint detected between some sequences does not have a counterpart in others. Consequently, all gap characters are treated as unknown nucleotides rather than as deletions. Conserved sequences between two outgroup species, in this case the shark, *Heterodontus francisci*, and the bichir, *Polypterus senegalus*, are compared to pairs of ingroup sequences, in these cases all pairs of the zebrafish a/b clusters, the fugu a/b clusters and the Tilapia, *Oreochromis niloticus*, HoxAa cluster. The fraction of the conserved nucleotide positions between the two ingroup sequences is compared with a exponential decay model (Wagner et al. 2004), assuming that there is a constant rate of loss of conservation along the lineages. Finally these numbers are used for testing for differences in the rate of modification between the two ingroup lineages as described in (Wagner et al. 2004).

Background

Hox cluster duplication is followed by loss of some duplicated genes. The gene complement of various ray finned fish HoxA clusters has been described in previous papers (bichir: (Chiu et al. 2004); zebrafish: (Amores et al. 1998; Chiu et al. 2002); stripped bass: (Snell et al. 1999); fugu and puffer fish: (Amores et al. 2004); Tilapia: (Santini et al. 2003)). In addition the complete sequence of the shark HoxA cluster has been described (Chiu et al. 2002; Kim et al. 2000). Here we recapitulate the salient features of these results to set the stage for the detailed sequence analysis reported below (Fig. 2).

The HoxA cluster of the ancestral ray finned fish most likely had paralog group members 1 to 7, 9 to 11 and 13 and 14, i.e. 12 genes (Powers and Amemiya 2004). Prior to HoxA cluster duplication in the stem of teleost fishes the HoxA cluster further lost *Hoxa-6,* and *Hoxa-14* (Chiu et al. 2004; Powers and Amemiya 2004), and thus had a total of ten genes. After duplication the b-paralog cluster lost all the anterior and medial genes with the exception of *Hoxa-2b*. In contrast the HoxAa paralog cluster only lost one or two genes in different lineages. The fugu and pufferfish lineage both lost *Hoxa-7a* and zebrafish also *Hoxa-10a*, while tilapia and the stripped bass retained both of these genes. The fugu has the largest number of first order paralogs, namely all the AbdB related genes *Hoxa-9a/b, Hoxa-10a/b, Hoxa-11a/b* and *Hoxa-13a/b*, as well as *Hoxa-2a/b*. In zebrafish only three 1st order paralog HoxA genes are maintained: *Hoxa-9 a/b, Hoxa-11a/b* and *Hoxa-13 a/b*. The detailed sequence analysis presented here focuses on the Hox genes with first order paralogs, with a particular focus on *Hoxa-11* and *Hoxa-13*, and some results for *Hoxa-2, Hoxa-9* and *Hoxa-10*. The analysis of the divergence of non-coding sequences of fugu and zebrafish will be restricted to the intergenic regions between *Hoxa-9* and *Hoxa-13*, i.e. the segment of the cluster that contains most of the 1st order paralogs.

Inferences on the consequences of Hox gene duplication critically depend on the availability of close outgroup taxa. There is only a limited amount of data from basal actinopterygian HoxA clusters. Recently we described a complete HoxA cluster sequence from the most basal lineage of ray finned fishes, the bichir (Chiu et al. 2004), and two genes, *Hoxa-11* and *Hoxa-13*, from the American paddlefish, *Polyodon spathula* (Metscher et al. 2004), which is a member of the second most basal ray finned fish lineage (Bermis et al. 1997). Gene tree analysis of either taxon shows that these lineages arose prior to the origin of the paralog HoxA clusters found in the teleosts (Chiu et al. 2004; Metscher et al. 2004).

Results

Increased rate of synonymous and non-synonymous substitutions after Hox gene duplication

The amino acid substitution rates of the18 teleost Hox genes studied here were compared to the rate in either the paddlefish or the bichir using a maximum likelihood relative rate test as implemented in HYPHY99beta for MacOS. In every instance the estimated amino acid substitution rate in the teleost gene was higher than that in

bichir or paddlefish and in 15 cases the difference was statistically significant (Fig. 3, Table 1).

The higher rate of non-synonymous substitutions in duplicated Hox genes is expected assuming that duplicated genes have a certain degree of functional redundancy. Functional redundancy would imply that the non-synonymous rate would differentially increase over the synonymous rates, leading to an increased dN/dS ratio, ω. Given that the duplication may be as long as 320 Mio years ago (Vanderpoele et al. 2004), estimating the synonymous substitution rate is only possible with genes for which relatively close out groups are available. Table 2 gives the results for *Hoxa-11* and *Hoxa-13* since only for those genes we had a paddlefish ortholog at hand. Both dN and dS were estimated with a codon based maximum likelihood method (Codeml of PAML) (Yang 1997) and a distance based method (RRTree)(Robinson-Rechavi and Huchon 2000) and compared to that in paddlefish (Tables 2A and B). Both methods account for phylogenetic structure, which makes these methods more appropriate for the comparison of deeply divided lineages than three taxa relative rate tests. In addition RRTree estimates the transversion rate at fourfold degenerate sites (B). This measure of the synonymous substitution rate can be estimated over longer phylogenetic distances because it saturates more slowly than dS itself.

Unexpectedly, PAML-based tests for differences in ω between teleosts and paddlefish, unexpectedly, did not show significant differences (not shown) in spite of the large differences in amino acid substitution rates summarized in Table 1. The data in Tables 2A and B show that both, the synonymous as well as the non-synonymous substitution rate in the teleost lineage are increased by a factor 3 to 5, depending on the gene and the method of estimation. For *Hoxa-11* both dS as well as B4 are significantly higher in teleosts than in paddlefish according to RRTree. For *Hoxa-13* dS could not be estimated but B4 is significantly increased (differences in dS can not be tested in PAML). The increased synonymous substitution rate can be the result of changes in the base composition and/or codon bias or an increased mutation rate. Below we show that the first two explanations are not supported by the data.

Accelerated evolution after Hox cluster duplication is not caused by changes in compositional or codon bias

To address the issue why the rate of synonymous substitutions increased after Hox cluster duplications we consider three possibilities: a systematic shift in the base composition of the HoxA clusters, a change in codon bias and a general increase of mutation rate in all nuclear genes.

As shown in Tables 2A and B, the GC content of the teleost HoxA clusters and genes does not differ systematically from out groups. On average there is a slightly higher GC content of teleost genes but this difference is only due to perciform species, i.e. fugu, tiliapia and stipped bass, but not zebrafish (data not shown), which also has an increased *dS*. Hence there is no indication that the change in synonymous substitution rate is due to a shift in GC content. Similarly, the effective number of codons, N_c, a measure of codon bias (Wright 1990), is not systematically different between teleosts and outgroups like bichir, coelacanth, shark or paddlefish. There are a few genes which have changed their condon usage pattern, but they are single genes in single lineages, like fugu *Hoxa-2b*, and can not account for the overall increase in

synonymous substitution rate. Thus we conclude that the most likely explanation for the increased synonymous substitution rate is an increase in mutation rate.

In order to test whether the apparent increase in mutation rate is limited to HoxA genes or a genome wide pattern we examined another nuclear gene, *RAG1*. As mentioned above, the main limitation for detecting any rate differences in sequence evolution is the availability of close out-group taxa. We only found one gene with sufficient coverage of basal ray finned fish lineages in Genbank to estimate the synonymous rate differences between teleosts and basal ray finned fishes, *RAG1*.

Rate of RAG1 evolution in ray finned fishes

We obtained *RAG1* sequences from 12 species to compare the rate of evolution between teleost and basal ray finned fish lineages, in particular paddlefish and Amia. The rate of evolution was estimated and tested for difference with three methods: codon based PAML (Yang 1997), codon model of (Muse and Gaut 1994) as implemented in HYPHY.99beta, and tree based relative rate tests as implemented in RRTree (Robinson-Rechavi and Huchon 2000). RRTree tests for differences in non-synonymous substitution rates are all highly significant between Amia/Paddlefish and teleosts, as well as between Amia/Paddlefish and osteoglossomorphs. Hence teleost nuclear genes investigated here evolve at a higher non-synonymous rate than close outgroups.

The synonymous substitution rate could not be estimated with RRTree due to saturation and even B4, the four fold degenerate transversion rate, could only be estimated with exceedingly high standard deviations. We then performed an RRTree test with the 3rd codon positions. The nucleotide substitution rate at 3rd positions could be estimated with reasonable error variance and indicated no difference between teleost clades and Amia/Paddlefish. The synonymous substitution rates estimated by PAML are consistently higher in teleosts but no direct test for the significance of these estimates can be given in the PAML framework.

Finally we performed a local clock test based on the Muse and Gaut (1994) codon model for heterogeneity of dS rates. The results show that the teleosts are not different from a clock model, and neither is a group including the Amia and Paddlfish. Hence synonymous substitution rate is not distinguishable from a clock model in a clade which includes the basal rayfinned fishes Amia and Paddlefish as well as teleosts. A local clock test for non-synonymous substitutions, however, provides strong evidence for rate heterogeneity within this clade (LR=48.52, P=$8*10^{-8}$) but not within the teleosts clade. This is consistent with the results of tree based relative rate tests (see above).

In order to test the power of the local clock model to detect differences in the synonymous substitution rate we performed a local clock test with the *Hoxa-11* data. The results indicate significant rate heterogeneity at the level of inclusion predicted by the RRTree tests, i.e. significant heterogeneity in the clade including Paddlefish but not within the teleost gene clade. Hence the clock test for synonymous rates is powerful enough to detect synonymous rate heterogeneity in *Hoxa-11* sequences at that level of phylogenetic differentiation and with shorter sequences than the *RAG1* alignment. Hence it seems likely that the failure of detecting synonymous rate

heterogeneity for *RAG1* among the ray finned fishes is not due to lack of power. It rather seems that there is a much smaller synonymous substitution rate difference, than found in the Hox genes, between teleosts at the one side and Amia and paddlefish at the other side.

In summary, there is strong evidence for an increased non-synonymous substitution rate in teleosts for *RAG1*, but no evidence for an increase in the synonymous substitution rate in teleosts. While there are still questions about the power to detect differences in the synonymous substitution rate with taxa as diverged as paddlefish and zebrafish, the available results from local clock tests and the rate of 3rd codon positions argues against that possibility. Hence it is possible that the increased dS rate observed for HoxA genes is Hox gene specific rather than a genome wide phenomenon.

Asymmetry of divergence among 1st order paralogs

Among the zebrafish HoxA genes there are three 1st order paralog gene pairs: *Hoxa-13a/b, Hoxa-11a/b* and *Hoxa-9a/b* (Fig. 2). Among those genes only one, *Hoxa-13a*, has a significantly increased non-synonymous substitution rate compared to its 1st order paralog *Hoxa-13b* (Table 3). Fugu, on the other hand, has five 1st order HoxA paralog pairs: *Hoxa-13a/b, Hoxa-11a/b, Hoxa-10a/b, Hoxa-9a/b* and *Hoxa-2a/b* (Fig.2). Of these, three have a significantly increased rate of non-synonymous substitutions, namely *Hoxa-11b, Hoxa-10b* and *Hoxa-2b* (Table 3). Note that all of the accelerated genes are b-paralogs. Furthermore the estimated rates for all of the b-paralogs are higher than that of the a-paralogs even for those two cases where the differences are not statistically significant, *Hoxa-13a/b* and *Hoxa-9a/b*. The *a priori* probability that in five instances paralogs from the same cluster has a higher estimated rate than the other is $0.5^5 = 0.03125<5\%$. Hence it is possible that in fugu the b-paralogs are diverging at a systematically higher rate than the a-paralogs, rather than paralogs randomly sampled from both paralog clusters.

Asymmetry of divergence of conserved non-coding sequences

The majority of 1st order paralog gene pairs are found in the 5' segment of the HoxA gene cluster. To compare the rate of non-coding sequence evolution after cluster duplication we focused on the inter-genic sequences 3' of *Hoxa-13* and 5' of *Hoxa-9*. In this analysis we also included the published sequence of the Tilapia HoxA-a cluster (Santini et al. 2003). A global alignment of the 5'HoxA clusters of shark and bichir identified blocks of conserved nucleotides as described in Prohaska et al. (2004). Then we identified the corresponding blocks of conservation in the five teleost HoxA clusters from zebrafish, fugu and tilapia and subjected them to a relative rate test as described in (Wagner et al. 2004). The results indicate that three of the duplicated 5'HoxA clusters are indistinguishable with respect to the rate of modification of conserved non-coding sequences among the clusters: the zebrafish HoxA-b, Tilapia HoxA-a and Fugu HoxA-a clusters (Table 4). Two clusters show a significant increase in the rate of modification of conserved non-coding sequences: zebrafish HoxA-a and fugu HoxA-b clusters; but these are indistinguishable from each other.

The results of the asymmetric modification of conserved non-coding sequences show an intriguing pattern of congruence with the divergence of the coding sequences. In

zebrafish the only gene with asymmetric divergence is a HoxA-a paralog, *Hoxa-13a* and the non-coding sequences of the HoxA-a cluster are diverging at a higher rate than that of the zebrafish HoxA-b cluster. There is however, a possible confounding factor in this comparison since it is also the HoxA-a cluster that lost its *Hoxa-10a* paralog and thus may lose non-coding sequence conservation at a higher rate than HoxA-b cluster because of that.

In fugu, however, the pattern is clearer. All five b-paralogs are estimated to have increased rates of non-synonymous substitutions (although only three of them are statistically significant, Table 3) and also the non-coding sequences of the HoxA-b cluster diverge faster than the fugu HoxA-a cluster. Notably the non-coding sequences of the fugu HoxA-b cluster also diverge faster than the zebrafish HoxA-b cluster. The latter observation is significant because both, the zebrafish and the Fugu HoxA-b clusters have the same gene inventory (Fig. 2). Hence the increased rate of non-coding sequence evoolution in the fugu HoxA-b cluster cannot be explained by differences in the number of genes retained. Thus in the fugu we observe a concerted asymmetric divergence of both the coding as well as the non-coding sequences of the 5'HoxA-b clusters after duplication. Furthermore this patter is distinctly different from that observed in the zebrafish, where the HoxA-a cluster divergences at a higher rate. Hence the majority of the asymmetric divergence occurred after the split of the zebrafish and the fugu lineage, rather than immediately after Hox cluster duplication.

In order to examine whether the non-coding sequence evolution rate increase is uniform along the Hox cluster we performed relative rate tests on each of the intergenic regions independently (Fig. 4). In the zebrafish HoxAa cluster the higher rate of divergence of the a-paralog is focused on the intergenic regions between *Hoxa-13* and *Hoxa-11* as well as between *Hoxa-11* and *Hoxa-10*. These two regions are 3' of *Hoxa-13*, which has an accelerated rate of non-synonymous divergence and 5' of *Hoxa-10*, which in the HoxAa cluster is a pseudogene. In fugu, the only significantly increased rate of non-coding divergence is between *Hoxa-10* and *Hoxa-9,* where *Hoxa-10* has an accelerated rate of evolution. Hence, the rate acceleration is differential among different ingergenic regions and could provide hints at the molecular level of regulatory changes following Hox cluster duplications.

Discussion

Rate of sequence evolution in teleosts

For all genes analyzed in this paper, including the non-Hox gene *RAG1*, the rate of non-synonymous substitutions is higher in teleosts than in close out groups, i.e. basal bony fish lineages like paddlefish and bichir. This finding is consistent with at least two previously published reports. Robinson-Rechavi and Laudet (2001) reported that most nuclear genes evolve quicker in teleosts than in mammals, independent of the presence of paralog genes, i.e. teleost genes evolve quicker regardless of whether there is potential for genetic redundancy or not. Krieger and Fuerst (2002) reported a smaller rate of amino acid evolution in sturgeons compared to teleost genes, in accordance to the results reported here. The interpretation of these authors, however, was that the rate in sturgeons is depressed, but this feature is shared with all basal gnathostome lineages, including bichir and Latimeria. Hence it is more parsimonious

to assume that the high rate of amino acid sequence evolution in teleosts is derived, rather than the low rate in bichir and paddlefish.

In teleost Hox genes the increased rate of non-synonymous substitutions is not necessarily indicative of relaxed purifying selection or directional selection. Other studies have established that recently duplicated genes have a significantly increased dN/dS rate ratio ω indicating either relaxed selection or directional selection (Conant and Wagner 2003; Kondrashov et al. 2002; Lynch and Conery 2000). The failure to detect an increased ω in teleost Hox genes may be due to two factors. First, the duplication of teleost genes are likely associated with a genome duplication which may have occurred as long as 320 Mio years ago (Vanderpoele et al. 2004). Lynch and Connery (2000) have shown that older duplicates return to lower ω values than recent duplicates. On the other hand, we report evidence that the synonymous substitution rate, dS, is also substantially increased. An increased rate of dS makes it harder to detect relaxed stabilizing selection.

Not many studies have addressed the questions whether synonymous substitution rates differ between teleosts and other bony fish. Merrit and Quattro studied the evolution of cytosolic malate dehydrogenase genes and reported an increased rate of evolution of teleost genes compared to sturgeon and mammalian genes. The authors report an increased dN/dS ratio and even directional selection in one lineage, i.e. dN/dS>1. The average number of synonymous changes in the teleost lineages (111.75), however, is similar to that in the sturgeon lineage (99) and not comparable with the roughly threefold increase we found in the HoxA genes in this study. We are not aware of any study that has investigated dS for other teleost nuclear genes and compared it with basal actinopterygians. Our attempt to test whether the non-Hox gene *RAG1* also has increased dS were ambiguous, but the balance of evidence suggests that dS for *RAG1* is not increased among teleosts, at least not as much as in HoxA genes. Hence, it is possible that the increased rate of synonymous substitutions is a phenomenon restricted to duplicated Hox clusters or may be even only the HoxA clusters of teleosts.

Of the possible explanations for an increased rate of synonymous substitutions we could eliminate two: changes in base composition and changes in the degree of codon bias. We thus conclude that the most likely explanation for the increased dS is an increased mutation rate but also note that this conclusion is only supported by the elimination of known alternatives and not by direct evidence. It would be desirable to compare the mutation rates of homologous genes in teleost and non-teletost fishes.

Asymmetrical divergence of paralog genes

There are a total of eight 1st order HoxA paralog gene pairs, three gene pairs in zebrafish and five in fugu. Four of those paralog pairs (50%) diverge at a rate that is statistically distinguishable at the 5% level. This fraction is comparable of that reported by Zhang et al (Zhang et al. 2003) for recent duplicates in the human genome (60%). G. Conant and A. Wagner (2003) report asymmetry in 20 to 30% of duplicated yeast, Drosophila and *Caenorhabditis elegans* genomes. Van de Peer and collaborators (2001) found that 50% of duplicated zebrafish genes have significant asymmetry in divergence. Kellis et al. (2004) report about 16% asymmetrically

diverging genes in yeast, although these numbers are not comparable because they did not apply rigorous statistical tests to detect asymmetry. In contrast, Kondrashov et al. (2002) only found significant asymmetry in two out of 49 (4%) recently duplicated mammalian genes and two out of eleven (18%) in Drosophila. While there might be both biological as well as methodological reasons responsible for these discrepancies, the results show that the HoxA genes in zebrafish and fugu are at the upper end of the frequency of asymmetrically diverging genes families even though they are not recently duplicated like the mammalian genes reported by Zhang et al., (2003). This result of consistent with that of Van de Peer et al. (2001) who also found high levels of asymmetrical divergence in a sample of genes containing a large fraction of transcription factors.

In fugu the asymmetrical divergence of duplicated HoxA genes has two intriguing additional features. In all of the five 1st order paralog HoxA gene pairs the b-paralogs have a higher estimated rate of dN than the a-paralog (of which three are statistically significant), which by chance is less likely than 5%. It is thus possible that the asymmetric divergence of Hox genes on a cluster is not independent but is correlated perhaps because of the strong regulatory integration among the Hox genes. Acquisition of new functions by one Hox gene may increase the chance of functional recruitment of other genes on the same cluster. This conclusion is also supported by the fact that many secondary functions of AbdB Hox genes, like their role in paired appendage development and reproductive tract development involve multiple Hox genes, and often all the AbdB related genes on a cluster (Haack and Gruss 1993; Kobayashi and Behringer 2003; Taylor et al. 1997). These facts suggest that the set of AbdB related genes act like a co-evolutionary module (Schlosser 2002).

Furthermore, the conserved non-coding regions in the cluster segment 5' of Hoxa-9 (the AbdB related set of genes), also diverges asymmetrically with the fugu 5'HoxA-b cluster evolving faster. This suggest, that in fugu the HoxA-b cluster diverged faster with respect to protein coding sequences as well as putative cis-regulatory elements. In the zebrafish the HoxA-b cluster is neither accelerated in any of its protein coding genes nor with respect to its non-coding sequences. The situation in the zebrafish HoxA-a cluster is complicated due to a higher rate of gene loss than the fugu HoxA-a cluster. In any case the results show that the pattern of asymetric gene divergence differs between the zebrafish and the fugu lineage.

In both the zebrafish HoxAa cluster and the fugu HoxAb cluster the rate acceleration of non-coding sequence evolution is heterogenous along the cluster. In either case the intergenic segments that show accelerated evolution are localized in the vicinity of coding regions, which diverge asymmetrically (Fig. 4). Thus, it is likely that the regulatory changes that lead to the asymmetrical divergence of Hox genes are localized and thus specific to particular expression domains rather than distributed general along the whole paralog sequences.

The relationship between the rate of cis-regulatory element evolution and the rate of protein evolution has recently been investigated in nematodes (Castillo-Davis et al. 2004). These authors found no significant correlation between dN and the divergence of promoter proximal transcription factor binding sites if the regression is corrected for differences in dS, i.e. time of duplication and mutation rate. Hence the association reported here between putative cis-regulatory element evolution and Hox protein

divergence was not found in a genome wide study. The reason for this discrepancy might be that the mechanistic reason for protein divergence differs between different classes of proteins. Castillio-Davis and collaborators did not report whether the correlation between cis-regulatory element divergence and protein sequence divergence differs between different classes of proteins. For transcription factors differences in tissue expression is possibly associated with a different set of tissue specific protein interaction partners (Featherstone 2003). Hence it is mechanistically plausible that divergence in gene regulation, inferred from the divergence in putative cis-regulatory elements, and protein sequence evolution are more strongly correlated in transcription factor genes than in other genes.

It is noteworthy that the asymmetric divergence of paralog genes is different between the zebrafish and the fugu lineage. For instance, in zebrafish the *Hoxa-13a* paralog is diverging faster while in fugu it is the *Hoxa-13b* paralog. This implies that the faster divergence has happened after the split of the zebrafish and fugu lineages rather than immediately after the duplication event, suggesting that asymmetrical divergence is driven by forces that do not acting immediately after the duplication but take some time to mount and are random among lineages. This result could also explain why the frequency of asymmetrically diverging duplicated genes is very low among recently duplicated genes (4 to 18%, (Kondrashov et al. 2002)). An interpretation of this pattern is that asymmetrical divergence is indicative of a novel function by the faster diverging gene (this possibility is considered further below with the zebrafish *Hoxa-13* genes). The acquisition of a novel function is probably only possible after a considerable period of random divergence due to relaxed stabilizing selection. This scenario would explain the delay after duplication in the development of asymmetric divergence.

Conclusions:

Based on the results reported here we propose a two stage model of Hox gene evolution following Hox cluster duplication consistent with the idea that duplicated Hox genes can contribute to physiological or developmental innovations (Malaga-Trillo and Meyer 2001; Ohno 1970; Wagner et al. 2003).

1. Immediately following the Hox cluster duplication there was a phase of increased random divergence including gene loss and increased mutation rates.
2. Some time after the duplication one paralog of a subset of the duplicated genes acquires novel functions and then diverges faster than the other paralog.

Evidence for a phase of increased random divergence following duplication comes from the increased rate of non-synonymous substitutions in teleosts. The Hox genes share this feature with other nuclear genes. In addition, however, there is also evidence for an increased mutation rate of Hox genes, not seen in other nuclear genes. Hence, in Hox genes at least, the increased coding sequence divergence is not alone caused by relaxed stabilizing selection due to genetic redundancy, but probably also by enhanced mutation rate.

Strongly asymmetric divergence of first order paralogs has been considered as evidence for the acquisition of new functions by one paralog and the maintenance of the ancestral function by the other paralog. The asymmetry arises to a large degree after the split of zebrafish and fugu lineages and differs between these two lineages. This feature is also suggestive of the acquisition of a new function, since it is likely that recruitment in a new function requires some time.

Acknowledgements: we thank Dr Gavin Conant for reading an earlier version of this manuscript. The financial support by NSF grant INB-0321470 to GPW and the DFG Bioinformatics Initiative BIZ-6/1-2 to PFS is gratefully acknowledged.

Table 1: comparison of amino acid substitution rate of teleost genes compared to a non-duplicated outgroup. dA stands for the maximum likelihood estimate of the amino acid substitution rate as calculated by HYPHY under the Dayhoff substitution model (Muse and Gaut 1994). The subscript T indicates the rate of the teleost gene lineage and O stands for that of the outgroup lineage. Note that all ratios dA_T/dA_O are larger than 1, indicating that the rate in the teleost lineage is estimated to be higher than in the outgroup lineage. Fifteen of the eighteen comparisons (83%) are statistically significant.

Gene	dA_T	dA_T/dA_O	P	outgroup
Dr-Hoxa-2b	0.060	1.598	0.538 n.s.	bichir
Tr-Hoxa-2a	0.097	1.898	0.289 n.s.	bichir
Tr-Hoxa-2b	0.198	6.821	0.001 **	bichir
Dr-Hoxa-9a	0.546	2.045	0.003 **	bichir
Dr-Hoxa-9b	0.367	1.872	0.029 *	bichir
Tr-Hoxa-9a	0.606	2.304	0.0004 ***	bichir
Tr-Hoxa-9b	0.529	2.726	0.0003 ***	bichir
Dr-Hoxa-10b	0.285	2.500	0.047 *	bichir
Tr-Hoxa-10a	0.308	2.053	0.087 n.s.	bichir
Tr-Hoxa-10b	0.473	6.088	0.0005 ***	bichir
Dr-Hoxa-11a	0.175	4.768	0.0002 ***	paddlefish
Dr-Hoxa-11b	0.147	4.153	0.0009 ***	paddlefish
Tr-Hoxa-11a	0.154	3.775	0.0012 **	paddlefish
Tr-Hoxa-11b	0.235	6.546	$9.3\ 10^{-7}$ ****	paddlefish
Dr-Hoxa-13a	0.202	4.810	$2.8\ 10^{-6}$ ****	paddlefish
Dr-Hoxa-13b	0.157	4.448	$6.9\ 10^{-5}$ ****	paddlefish
Tr-Hoxa-13a	0.131	2.433	0.0086 **	paddlefish
Tr-Hoxa-13b	0.207	2.875	0.0002 ***	paddlefish

Table 2: comparison of nucleotide substitution rates, base composition and codon bias between teleost Hox genes and paddlefish. The PAML rate estimates for the teleost are averages over the terminal branches, while the RRTree estimates are averages over the whole clade. For *Hoxa-13* the synonymous rate could not be estimated most likely because of saturation. Note that both the non-synonymous as well as the synonymous rates are increased in the duplicated teleost genes. The increase in synonymous substitution rate could neither be explained by a shift in GC content nor by a chance in codon bias. dN = non-synonymous substitution rate; dS = synonymous substitution rate; B4 = transversion rate at fourfold degenerate sites. GC = fraction of G and C in the coding sequence; GC3 = fraction of G and C in third codon positions; Nc = effective number of codons (high Nc indicates low codon bias; (Wright 1990)).

A) *Hoxa-11*

	RRTree			PAML				
Hoxa-11	dN	dS	B4	dN	dS	GC	GC3	Nc
Teleosts	0.116	1.554	0.892	0.066	1.074	0.53	0.63	57.8
Polyodon	0.032	0.508	0.195	0.012	0.353	0.49	0.59	56.6
Factor	3.6	3.1	4.5	5.5	3.0	1.1	1.07	1.02
P	10^{-7}	$3.6*10^{-5}$	$9.6*10^{-3}$					

B) *Hoxa-13*

	RRTree			PAML				
Hoxa-13	dN	dS	B4	dN	dS	GC	GC3	Nc
Teleosts	0.123	---	0.91	0.0621	0.970	0.52	0.56	57.1
Polyodon	0.037	----	0.33	0.0103	0.129	0.50	0.48	58.4
Factor	3.3	---	2.7	6.0	7.5	1.0	1.2	1.00
P	10^{-7}	sat	$4.1*10^{-4}$					

C) *RAG1*

	RRTree		PAML				
RAG1	dN	B4	dN	dS	GC	GC3s	Nc
Teleosts	0.139	5.88#	0.182	1.79	0.52	57.5	54.1
(*Polyodon, Amia*)	0.093	4.4	0.087	0.548	0.51	56.5	53.6
Factor	1.5	1.2	2.1	3.26	1.0	1.0	1.0
P	$9*10^{-5}$	1.00#					

#) Saturation, test is not working

Table 3: comparison of non-synonymous substitutions rates of paralog gene pairs. A maximum likelihood relative rate test was performed using the codon model of Muse and Gaut (1994) as implemented in HYPHY. The outgroup for *Hoxa-2, Hoxa-10,* and *Hoxa-11* was the corresponding human gene, for *Hoxa-9* and *Hoxa-13* the outgroup was *Xenopus tropicalis*. LR stands for likelihood ratio.

Gene	dN of paralog-a	dN of paralog-b	LR	P
Dr Hoxa-9	0.1497	**0.2082**	0.0820	0.775
Dr Hoxa-11	**0.0876**	0.0837	0.0176	0.894
Dr Hoxa-13	**0.1256**	0.0613	5.0095	**0.025**
Tr Hoxa-2	0.0796	**0.1189**	4.2991	**0.038**
Tr Hoxa-9	0.1213	**0.1899**	0.1389	0.709
Tr Hoxa-10	0.1683	**0.1935**	9.3324	**0.002**
Tr Hoxa-11	0.0805	**0.1544**	4.7850	**0.028**
Tr Hoxa-13	0.0796	**0.1098**	1.3691	0.242

Table 4: comparison of the rate of evolution of conserved non-coding regions in the inter-genic regions between *Hoxa-13* and *Hoxa-9* of the zebrafish and fugu HoxA clusters according to the method described in Wagner et al., (2004). In this method two distant outgroups are compared, in this case shark and bichir, to identify conserved non-coding sequences. Then the orthologous sequence elements are identified in the two ingroup sequences, in this case the two paralog clusters in zebrafish and fugu and the amount of sequence conservation compared. The degree of modification is here measured by the rate of retained conserved nucleotide positions r. A modified z'-statistic is calculated to account for stochastic dependence among the nucleotide positions. This test allows to assess whether the retention rates of conserved nucleotide positions is statistically different. The results show that the zebrafish Hox cluster segment DrAa13-9, i.e. the a-paralog, retains significantly fewer conserved nucleotide positions (about 20%) than its zebrafish paralog DrAb13-9 as well its orthologs in fugu and tilapia, TrAa13-9 and OnAa13-9. The results also show that the fugu TrAb13-9, the b-paralog, evolves faster in its putative cis-regulatory elements than its paralog TrAa13-9 and its zebrafish ortholog DrAb13-9.

Cluster 1	Cluster 2	r(cluster 1)	r(cluster 2)	z'	P
DrAa13-9	DrAb13-9	**0.200**	0.260	1.685	**0.046**
DrAa13-9	TrAa13-9	**0.199**	0.282	2.838	**0.002**
DrAa13-9	OnAa13-9	**0.197**	0.286	3.056	**0.001**
TrAa13-9	OnAa13-9	0.286	0.292	0.302	0.375
TrAa13-9	**TrAb13-9**	0.295	**0.206**	2.589	**0.005**
DrAb13-9	**TrAb13-9**	0.264	**0.203**	1.873	**0.031**
DrAb13-9	TrAa13-9	0.263	0.289	0.826	0.203

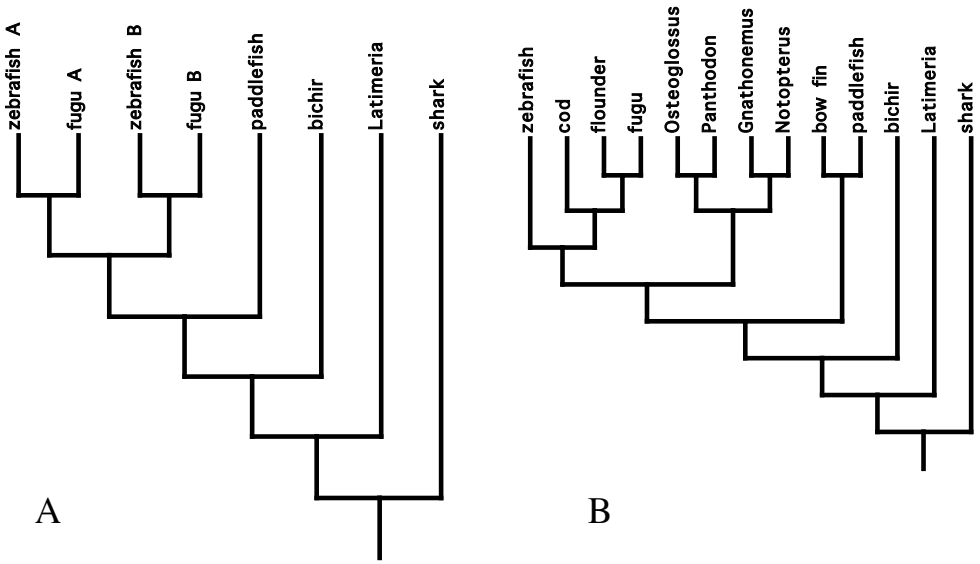
Figure captions

Figure 1: phylogenetic relationships among the gene sequences analyzed in this paper. A) gene tree of the HoxA genes analyzed. Bichir is considered the most basal ray finned fish lineage. Paddlefish together with the sturgeons belongs to the second most basal ray finned fish lineage (Inoue et al. 2003), and the two paralogs of the HoxA clusters in fugu and zebrafish derive from the same duplication event (Amores et al. 2004). B) phylogenetic relationships among the RAG1 genes analyzed in this paper. The bowfin, Amia calva, and paddlefish are considered as belonging to a clade according to the results of Inoue et al. (2003). The relationships of the osteoglossomorph sequences (Osteroglossus, Panthodon, Gnathonbemus and Notopterus) are based on (Guo-Qing and Wilson 1996).

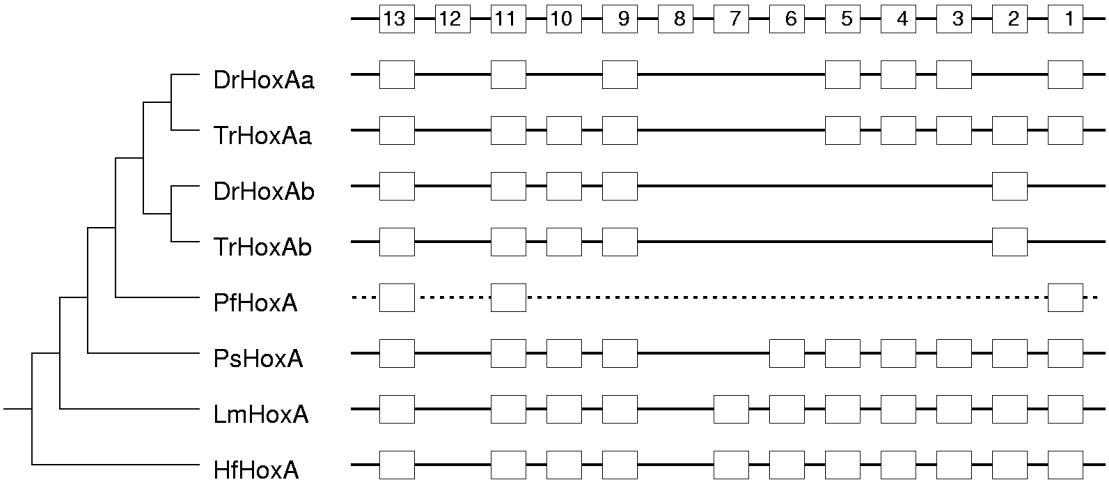
Figure 2: the gene inventory of the HoxA clusters analyzed in this paper. The shark, *Heterodontus francisci*, HoxA cluster, HfHoxA, has been described in Kim et al., (2000) and Chiu et al., (2002) as HoxM cluster. The coelacanth, *Latimeria menadoensis*, HoxA cluster, LmHoxA, is described in Powers et al., (2004), and the bichir, *Polypterus senegalus*, cluster, PsHoxA, has been described in Chiu et al., (2004). The paddlefish Hox genes are only known from three partial HoxA gene sequences: *Hoxa-13,* and *Hoxa-11* are represented by almost complete coding sequences including the intron (Metscher et al. 2004) and *Hoxa-1* by a smaller fragment (J. McClintock, C. Jozefowicz, and V. E. Prince, AY188522). The zebrafish HoxA clusters, DrHoxAa/b, are based on the complete Hox cluster sequence described in Chiu et al. (2002), and the fugu sequences, TrHoxAa/b, are from the fugu genome sequence versions 2.0 and 3.

Figure 3: amino acid sequence evolution of Hoxa-11 (A) and Hoxa-13 (B) duplicated gene is accelerated compared to the non-duplicated genes of paddlefish, bichir and coelacanth.

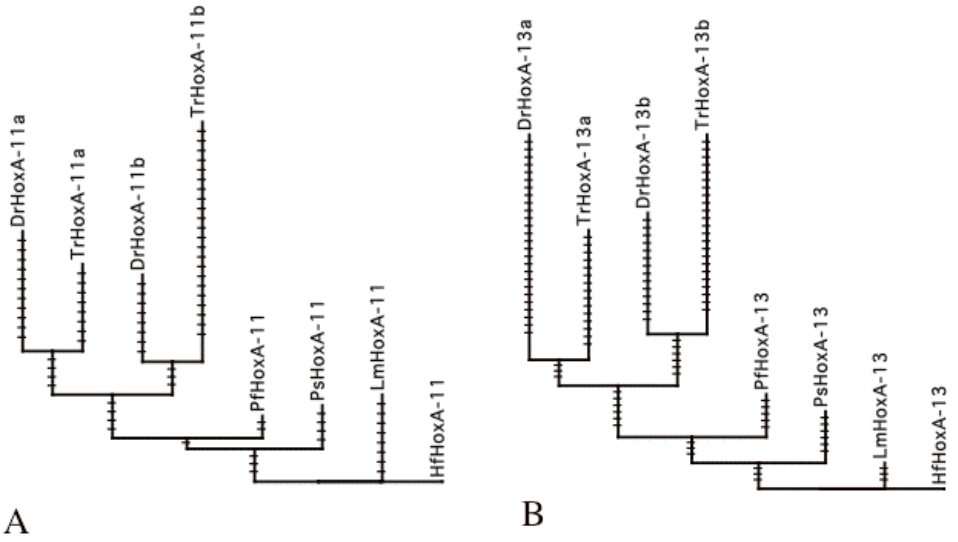
Figure 4: spatial distribution of sequence divergence along the 5' segment of the duplicated teleost HoxA clusters. The open boxes symbolize the coding regions (ignoring introns) and the lines the intergenic sequences (not to scale). The gray boxes in the intergenic regions give the difference in the fraction of conserved non-coding sequence positions modified in one of the two paralogs: (1-r(HoxAa) – (1-r(HoxAb)). If the difference is >0, i.e. if the HoxAa cluster is more modified than the HoxAb cluster the bar is on top of the line, if the HoxAb is more modified, the bar is drawn below the line. In the coding regions the gray bars symbolize the difference in the rate of non-synonymous substitutions (Table 3). Symbols: * P<5%, ** P<2.5%, *** P<1%; n.s. P>5%.



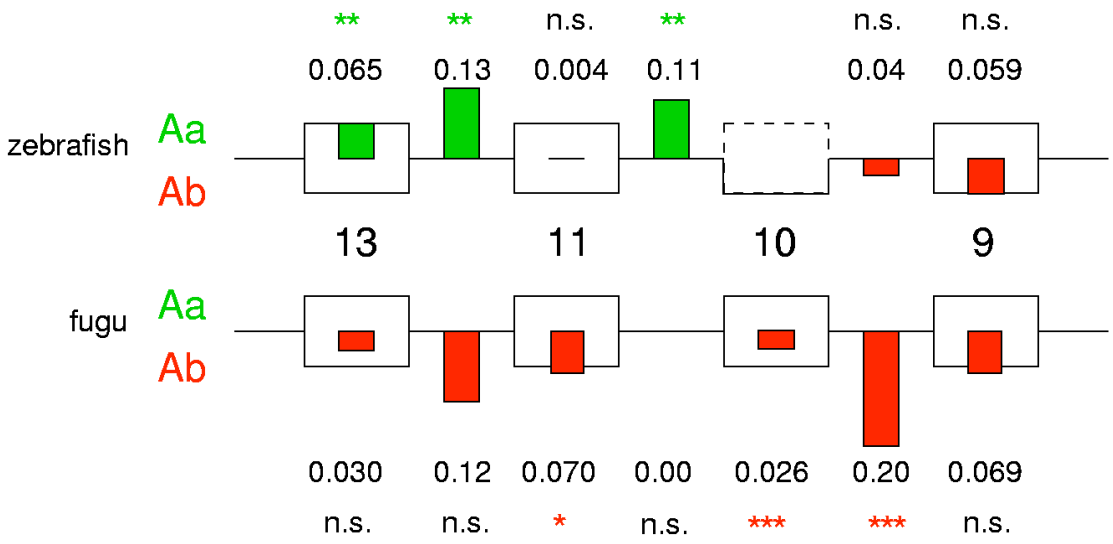
Wagner et al. Figure 1



Wagner et al., Figure 2



Wagner et al., Figure 3



Wagner et al., Figure 4

References:

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J. Mol. Biol. 215:403-410

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

Amores A, Suzuki T, Yan Y-L, Pomroy 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

Bermis WE, Findeis EK, Grande L (1997) An overview of Acipenseriformes. Environ. Biol. Fish. 48:25-71

Castillo-Davis CI, Kondrashov FA, Hartl DL, Kulathinal RJ (2004) The functional genomic distribution of protein divergence in two animal phyla: coevolution, genomic conflict, and constraint. Genome Research 14:802-811

Chiu C-h, Amemiya C, Dewar K, Kim C-b, Ruddle F, Wagner GP (2002) Molecular evolution of the HoxA cluster in three major gnathostome lineages. PNAS 99:5492-5497

Chiu C-h, Dewar K, Wagner GP, Takahashi K, Ruddle F, Ledje C, Bartsch P, Scemama J-L, 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 Research 14:11-17

Conant GC, Wagner A (2003) Asymmetric sequence divergence of duplicate genes. Genome Research 13:2052-2058

Dehal P, et a (2002) The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins. Science 298:2157-2167

Featherstone M (2003) HOX proteins and their co-factors in trancriptional regulation. In: Lufkin T (ed) Murine homeobox gene control of embryonic patterning and organogenesis. Elsevier, Amsterdam, NL, p 1-42

Ferrier DK, Minguillón C, Holland PWH, Garcia-Fernández J (2000) The amphioxus Hox cluster: deuterostome posterior flexibility and Hox14. Evol. Dev. 2:284-293

Fried C, Prohaska SJ, Stadler PF (2003) Independent Hox-cluster duplications in Lampreys. J. Exp. Zool. (Mol. Dev. Evol.) Part B 299:18-25

Fried C, Prohaska SJ, Stadler PF (2004) Exclusion of repetitive DNA elements from gnathostome Hox clusters. J Exp Zool Part B (Mol Dev Evol) 302B:165-173

Guo-Qing L, Wilson MVH (1996) Phylogeny of Osteoglossomorpha. In: Stiassny M, Parenti L, Johnson GD (eds) Interrelationships of Fishes. Academic Press, San Diego, CA, p 163-174

Haack H, Gruss P (1993) The Establishment of Murine Hox-1 Expression Domains during Patterning of the Limb. Developmental Biology 157:410-422

Holland PW, Garcia-Fernandez J (1996) Hox genes and chordate evolution. Developmental Biology 173:382-395

Holland PWH, Garcia-Fernández J, Williams NA, Sidow A (1994) Gene duplication and the origins of vertebrate development. Development (Suppl.):125-133

Inoue JG, Miya M, Tsukamoto K, Nishida M (2003) Basal actinopterygian realtionships: a mitogenomic perspective on the phylogeny of the "ancient fish." Mol Phyl. Evol 26:110-120

Ji FY, Liu JD, Yi MS, Huang L, Zhou F, Yu QX (2002) Chromosomal localization of rice filed eel Hox genes by PRINS. Acta Genetica Sinica 29:612-615
Kappen C, Schugart K, Ruddle FH (1989) Two steps in the evolution of Antennapedia-class vertebrate homeobox genes. PNAS 86:5459-5463
Kellis M, Birren BW, Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. Nature 428:617-624
Kim C-B, Amemiya C, Bailey W, Kawasaki K, Mezey J, MillerII W, Minoshima S, Shimizu N, Wagner GP, Ruddle F (2000) Hox cluster genomics in the horn shark, Heterodontus francisci. PNAS 97:1655-1660
Kobayashi A, Behringer RR (2003) Developmental genetics of the female reproductive tract in mammals. Nature Rev Genet. 4:969-980
Kondrashov FA, Rogozin IB, Wolf YI, E.V K (2002) Selection in the evolution of gene duplications. Genome Biology 3:REASEARCH0008
Krieger J, Fuerst PA (2002) Evidence for a slowed rate of molecular evolution in the order Acipenseriformes. Mol. Biol. Evol. 19:891-897
Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151-1155
Malaga-Trillo E, Meyer A (2001) Genome duplication and accelerated evolution of Hox genes and cluster architecture in teleost fishes. Amer. Zool. 41:676-686
Martinez P, Amemiya CT (2002) Genomics of the HOX gene cluster. Comp. Biochem. Physiol. Part B 133:571-580
Metscher BD, Takahashi K, Crow K, Amemiya C, Nonaka DF, Wagner GP (2004) Expression of Hoxa-11 and Hoxa-13 in the pectoral fin of a basal rayfinned fish, Polyodon spathula: implications for the origin of tetrapod limbs. Evolution and Development revision
Meyer A, Malaga-Trillo E (1999) Vertebrate genomics: more fishy tales about Hox genes. Curr. Biol. 9:210-213
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
Misof BY, Wagner GP (1996) Evidence for four Hox clusters in the Killifish, Fundulus heteroclitus (Teleostei). Mol. Phylogen. Evol. 5:309-322
Morgenstern B (1999) DIALIGN 2:improvement of the segment-to-segment approach to multiple sequence alignment. Bioinformatics 15:21-218
Muse SV, Gaut BS (1994) A likelihood approach for comparing synonymous and non-synonymous nucleotide sequences and its applications. Mol. Biol. Evol. 11:715-724
Naruse K, Fukamachi S, Mitani H, Kondo M, Matsuoka T, Kondo S, Hanmura 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
Naruse K, Tanaka M, Mita K, Shima A, Postleithwait J, Mitani H (2004) A medaka gene map: the trace of ancestral vertebrate proto-chromosomes revewaled by compariative gene mapping. Genome Research 14:820-828
Ohno S (1970) Evolution by Gene Duplication. Springer Verl., New York
Pavell AM, Stellwag EJ (1994) Survey of Hox-like genes in the teleost Morone saxatilis: implications for evolution of the Hox gene family. Mol. Marine Biol. Biotech. 3:149-157

Peer YVd, Taylor JS, Braasch I, Meyer A (2001) The ghost of selection past: rates of evolution and functinal divergence of anciently duplicated genes. J. Mol. Evol. 53:436-446
Powers TP, Amemiya CT (2004) Evidence for a Hox14 paralog group in vertebrates. Curr. Biol. 14:R183-184
Prince V, Pickett FB (2002) Splitting pairs: the diverging fates of duplicated genes. Nature Reviews Genetics 3:827-837
Prohaska SJ, Fried C, Flamm C, Wagner GP, Stadler PF (2004) Surveying phylogenetic footprints in large gene clusters: application to Hox cluster duplications. Mol. Phylo. Evol. 31:581-604
Prohaska SJ, Stadler PF (2004) The duplication of the Hox gene clusters in teleost fishes. Theory in Bioscience 123:89-110
Robertson-Rechavi M, Laudet V (2001) Evolutionary rates of duplicated genes in fish and mammals. Mol. Biol. Evol. 18:681-683
Robinson-Rechavi M, Huchon D (2000) RRTree: relative-rate tests between groups of sequences on a phylogenetic tree. Bioinformatics 16:296-297
Ruddle FH, Bartels JL, Bentley KL, Kappen C, Murta MT, Pendelton JW (1994a) Evolution of Hox genes. Annu. Rev. Genet. 28:423-442
Ruddle FH, Bentley KL, Murtha MT, Risch N (1994b) Gene loss and gain in the evolution of the vertebrates. Development (Suppl):155-161
Ruvkun G, Hobert O (1998) The taxonomy of developmental control in Caenorhabditis elegans. Science 282:2033-2041
Santini S, Boore JL, Meyer A (2003) Evolutionary conservation of regulatory elements in vertebrate Hox gene clusters. Genome Res. 13:1111-1122
Schlosser G (2002) Modularity and the units of evolution. Theory in Bioscience 121:1-80
Snell EA, Scemama JL, Stellwag EJ (1999) Genomic organization of the Hoxa-4-Hoxa10 region from Morone saxatilis: implications for Hox gene evolution among vertebrates. J. Exp. Zool. (Mol. Dev. Evol.) 285:41-49
Taylor HS, Vanden HG, Igarashi PA (1997) A conserved Hox axis in the mouse and human reproductive system: late establishment and persistent expression of the Hoxa cluster genes. Biol Reprod 57:1338-1345
Taylor JS, Braasch I, Frickey T, Meyer A, Peer YVd (2003) Genome duplication, a trait shared by 22,000 species of ray-finned fish. Genome Research 13:382-390
Vanderpoele K, Vos WD, Taylor JS, Meyer A, Peer YVd (2004) Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. PNAS 101:1638-1643
Wagner GP, Amemiya C, Ruddle F (2003) Hox cluster duplications and the opportunity for evolutionary novelties. PNAS 100:14603-14606
Wagner GP, Fried C, Prohaska SJ, Stadler PF (2004) Divergence of conserved non-coding sequences: rate estimates and relative rate tests. Mol. Biol. Evol. in press
Wright F (1990) The 'effective number of codons' used in a gene. Gene 87:23-29
Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. CABIOS 13:555-556
Zhang P, Gu Z, Li WH (2003) Different evolutionary patterns between young duplicate genes in the human genome. Genome Biology 4:R56