

## Genome Duplications and Accelerated Evolution of *Hox* Genes and Cluster Architecture in Teleost Fishes<sup>1</sup>

EDWARD MÁLAGA-TRILLO\* AND AXEL MEYER<sup>2\*</sup>†

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

†DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94538

**SYNOPSIS.** The early origin of four vertebrate *Hox* gene clusters during the evolution of gnathostomes was likely caused by two consecutive duplications of the entire genome and the subsequent loss of individual genes. The presumed conserved and important roles of these genes in tetrapods during development led to the general assumption that *Hox* cluster architecture had remained unchanged since the last common ancestor of all jawed vertebrates. But recent data from teleost fishes reveals that this is not the case. Here, we present an analysis of the evolution of vertebrate *Hox* genes and clusters, with emphasis on the differences between the *Hox* A clusters of fish (actinopterygian) and tetrapod (sarcopterygian) lineages. In contrast to the general conservation of genomic architecture and gene sequence observed in sarcopterygians, the evolutionary history of actinopterygian *Hox* clusters likely includes an additional (third) genome duplication that initially increased the number of clusters from four to eight. We document, for the first time, higher rates of gene loss and gene sequence evolution in the *Hox* genes of fishes compared to those of land vertebrates. These two observations might suggest that two different molecular evolutionary strategies exist in the two major vertebrate lineages. Preliminary data from the African cichlid fish *Oreochromis niloticus* compared to those of the pufferfish and zebrafish reveal important differences in *Hox* cluster architecture among fishes and, together with genetic mapping data from Medaka, indicate that the third genome duplication was not zebrafish-specific, but probably occurred early in the history of fishes. Each descending fish lineage that has been characterized so far, distinctively modified its *Hox* cluster architecture through independent secondary losses. This variation is related to the large body plan differences observed among fishes, such as the loss of entire sets of appendages and ribs in some lineages.

### INTRODUCTION

The understanding of the interconnectedness of developmental and evolutionary biology advanced much since the finding that the process of pattern formation in most—if not all—metazoan phyla is regulated by homologous sets of highly conserved developmental control genes (Akam, 1989; De Robertis, 1997; Slack *et al.*, 1993). Among these, *Hox* genes have become a paradigm for researchers who try to understand the generation of novel morphologies and the evolution of body plans, mainly because of their important role in the speci-

fication of the embryonic body axis (Lewis, 1978 and papers in this symposium volume). *Hox* genes are characterized by the presence of a 183 bp DNA sequence motif, the homeobox, which encodes a conserved DNA binding structure, the homeodomain (*e.g.*, reviewed in Gehring, 1998). Within the large homeobox gene superfamily, *Hox* genes are a subset defined by their arrangement in genomic clusters, and by their colinearity, *i.e.*, the correlation between chromosomal organization, time of activation, and boundary of expression along the anterior-posterior (a-p) axis (*e.g.*, Krumlauf, 1994). Genes located progressively upstream in the complex are activated later and more posteriorly in development.

That the evolution of morphological diversity could have been facilitated by increasing levels of genetic complexity is

<sup>1</sup> From the Symposium *HOX Clusters and the Evolution of Morphology* presented at the Annual Meeting of the Society for Integrative and Comparative Biology, 4.

<sup>2</sup> Axel.Meyer@uni-konstanz.de.

based largely on the comparative study of vertebrate Hox clusters (e.g., Holland *et al.*, 1994; Ruddle *et al.*, 1994). It is thought that the ancestral Hox gene cluster arose from a single locus as a result of multiple tandem duplication events, and that subsequent rounds of entire genome duplication gave rise to the multiple clusters now found in higher vertebrates (Holland, 1997; Holland and Garcia-Fernandez, 1996; Meyer and Málaga-Trillo, 1999; Meyer and Scharl, 1999; Wittbrodt *et al.*, 1998). Protostome invertebrates and the deuterostome cephalochordate *Amphioxus* possess a single Hox cluster, while mammals have four clusters (A–D), each derived from a basic arrangement of 13 paralogous groups (McGinnis and Krumlauf, 1992). These clusters are likely to have arisen by whole-genome duplications, early in vertebrate evolution (although a less parsimonious scenario of individual chromosomal duplications cannot be completely ruled out). Representatives of basal lineages such as agnathan fishes are hence expected to have two to four Hox clusters (Meyer, 1996, 1998; Zhang and Nei, 1996).

Until recently, it had been assumed that the mammalian Hox cluster architecture of 39 genes in four clusters was a shared derived feature for all vertebrates. However, the characterization of seven Hox clusters in the zebrafish (Amores *et al.*, 1998) revealed that both the number of clusters and the number of genes per cluster are free to vary during evolution. Earlier work on pufferfish Hox genes (Aparicio *et al.*, 1997) already had suggested that at least the number of genes can vary, even if the number of Hox clusters in fish was originally thought to be fixed at four. Whether this variation is confined to one or a few teleost lineages, or whether it represents a more general trend in vertebrates is not known, although fragmentary evidence from other teleost fish already suggested that this duplication is not exclusive for the zebrafish (Misof and Wagner, 1996; Prince *et al.*, 1998; Naruse *et al.*, 2000).

Here we compare the Hox A clusters from some of the few fish model systems studied to date, with regard to gene content and DNA sequence variation. Included in

these analyses are some of our own preliminary data on the Hox gene architecture of the African cichlid fish *Oreochromis niloticus*. We also used available Hox sequences from some of the most important fish model systems, to examine their relationships and to analyze Hox gene/cluster evolution.

#### THE HOX CASE FOR A TELEOST-SPECIFIC GENOME DUPLICATION

Bony fish (class Pisces) are the most successful and “species-rich” group of vertebrates, comprising more than 25,000 species that encompass a huge spectrum of morphological variation. The division Teleostei within the Osteichthyes (bony fish) is made up of 38 orders (Nelson, 1994). Hox gene DNA sequences have been characterized from six of these orders: Cypriniformes (zebrafish), Tetraodontiformes (pufferfish), Cyprinodontiformes (medaka), Atheriniformes (killifish), Salmoniformes (salmon) and Perciformes (African cichlid and striped bass) (Amores *et al.*, 1998; Aparicio *et al.*, 1997; Kurosawa *et al.*, 1999; Misof *et al.*, 1996; Misof and Wagner, 1996; Pavell and Stellwag, 1994; Snell *et al.*, 1999). Even though these orders cover a good portion of the teleost radiation, the datasets differ largely in the power of the experimental approaches used to collect the DNA sequences and therefore do not always provide the complete and reliable information that is required for a comprehensive comparative analysis of Hox cluster variation. The most common method employed for the identification of Hox genes has been the sequencing of short PCR products amplified from genomic DNA with degenerate “universal” primers for the homeobox. This strategy permits the presence of different paralogous group members to be established, but it does not allow for precise identification of their cluster affiliation, due to the limited number of variable amino acid positions in the homeobox. Furthermore, gene absence cannot be unambiguously confirmed due to the inherent problems of PCR template competition when amplifying multiple targets with one primer pair. Moreover, in the absence of linkage data, it can be difficult to distinguish between duplicated paralogues and alleles at

one locus. The alternative to this strategy is the characterization of Hox-positive clones from genomic libraries which, although laborious and time-consuming, provides extensive DNA sequence data for coding and non-coding regions, as well as unambiguous information about gene assignments and cluster composition. Among teleost fishes, this sort of reliable genomic data is available only for zebrafish (Amores *et al.*, 1998), pufferfish (Aparicio *et al.*, 1997), striped bass (Snell *et al.*, 1999) and now partly for the African cichlid fish *Oreochromis niloticus* (Málaga-Trillo, Amores, McAndrew, Postlethwait and Meyer, unpublished data).

The initial evidence for Hox variation between fishes and tetrapods comes from a study by Aparicio and colleagues (1997), where a set of only 31 Hox genes arranged in four clusters were described for the pufferfish (*Fugu rubripes*). Three of the clusters could be assigned as orthologues of the mammalian complexes A, B and C, but the fourth cluster (originally designated as D) showed ambiguous similarities to both the A and the D clusters, and probably represents an extra cluster. The absence of eight genes relative to the 39 mammalian Hox genes suggested that gene loss might be a distinctive feature of Hox cluster evolution, at least in the very compact *Fugu* genome. The apparent reduced Hox gene complement of the pufferfish correlates well with its known small genome, and these genetic characteristics have been readily associated with reduced or lacking features of its morphology, such as the absence of ribs and pelvic bones (Aparicio *et al.*, 1997). However, there is no experimental evidence yet on whether the genetic and morphological simplification observed in this fish are linked by a causal relationship (Meyer, 1998; Meyer and Málaga-Trillo, 1999; Meyer and Schartl, 1999).

The existence of more than the expected four Hox clusters in fish was first reported by Prince *et al.* (1998), who described 42 Hox genes in the zebrafish. 39 of these genes were unequivocally assigned to four linkage groups that identified clusters A–D, homologous to their mammalian counterparts. However, three new genes mapped to

two additional linkage groups, indicating the presence of six rather than the expected four Hox gene clusters. Subsequent extensive characterization of zebrafish Hox genes by Amores *et al.* (1998) showed that the zebrafish genome contains at least 48 Hox genes arranged in at least seven Hox clusters. Phylogenetic analysis and genetic mapping revealed that the seven zebrafish clusters are orthologous to the mammalian ones and that they most likely arose through an additional genome duplication event experienced by the fish but not the tetrapod lineage. These Hox clusters were termed A $\alpha$ , A $\beta$ , B $\alpha$ , B $\beta$ , C $\alpha$ , C $\beta$  and D $\alpha$  (Amores *et al.*, 1998). The zebrafish data also provided an explanation for the ambiguous pufferfish D complex: upon sequence comparisons between zebrafish and pufferfish, it became clear that the problematic *Fugu* cluster is indeed an A $\beta$  and not a D cluster (Amores *et al.*, 1998; Aparicio, 2000). This suggested that the putative genome duplication of the zebrafish is shared by at least one other teleost lineage and that it is not specific to Cypriniform fishes, or the result of a specific polyploidization of the zebrafish genome (Meyer and Málaga-Trillo, 1999). Therefore, the initial estimate of four Hox clusters in the pufferfish is probably erroneous, and raises the possibility of the existence of up to four additional clusters (or their secondary loss).

Besides the detailed genomic studies in zebrafish and pufferfish, upcoming genomic data from the African cichlid fish *Oreochromis niloticus* (Málaga-Trillo, Amores, McAndrew, Postlethwait and Meyer, unpublished data) provide additional evolutionary information for the comparative analysis of teleost Hox gene clusters. Although the characterization of cichlid Hox genes has not been completed, there is already enough evidence for the presence of more than four Hox clusters (at least six) that also contain different sets of genes than those of other fish. The specific differences between the entire Hox complements of these fish are not the focus of this article, and will be discussed somewhere else. But for the time being, it is interesting to note that all genomic evidence from the above-mentioned systems (zebrafish, pufferfish,

cichlid) indicates that an actinopterygian ancestor already possessed eight Hox clusters and not four, as it had recently been hypothesized (Stellwag [1999] but see Meyer and Málaga-Trillo [1999] and Meyer and Schartl [1999]). Stellwag also argued that the known occurrence of polyploidy in the order Cypriniformes could in principle explain the seven zebrafish clusters as the result of a lineage-specific duplication. This assumption was based on the fact that, at the time, no more than four representatives from each paralogous groups had been isolated from other divergent teleost lineages such as Medaka and the striped bass (Kurosawa *et al.*, 1999; Pavell and Stellwag, 1994). However, the data for these two fish were generated in PCR surveys and therefore do not constitute conclusive evidence about the absence of specific clusters, *e.g.*, the zebrafish case (see Misof and Wagner [1996] *vs.* Amores *et al.* [1998]). In this regard, it should be emphasized that even when using a more thorough genomic approach, technical difficulties can limit the coverage of a Hox screening procedure, as appears to have been the case for the original pufferfish study by Aparicio *et al.* (1997), where only four Hox clusters were identified (Amores, personal communication). In addition, three new lines of evidence strengthen the idea that the zebrafish genome duplication is not limited to the fishes of the order Cypriniformes: 1) The presence of two A clusters in the pufferfish (Amores, personal communication), 2) The finding of more than four Hox clusters in cichlid fish (Málaga-Trillo, Amores, McAndrew, Postlethwait and Meyer, unpublished data), and 3) The recent mapping of Medaka *Hox* genes to seven linkage groups (Naruse *et al.*, 2000). In this context, the identification of five *Hox* group 9 sequences in the PCR-surveyed killifish (Misof and Wagner, 1996) also argues for the existence of more than four Hox clusters in this species, adding support to the thesis of an additional teleost-specific genome duplication that is not shared with the sarcopterygian lineage.

#### INSIGHTS FROM VARIATION IN HOX CLUSTER A

Here we focus our analysis on variation of vertebrate Hox A clusters, because they

illustrate the history of duplications and gene losses that are typical for Hox clusters. Currently, there are more points of comparison across the vertebrate phylogeny for the A clusters, including partial data on basal lineages such as the horned shark *Heterodontus francisci* (Kim *et al.*, 2000) and the Australian lungfish *Neoceratodus forsteri* (Longhurst and Joss, 1999). Most reconstructions of Hox cluster evolution are based on the assumption that sharing of genes is indicative of common descent, since independent gene loss is more parsimonious than the independent gaining of genes. It is therefore possible to reconstruct Hox ancestral states of vertebrates by using the Hox cluster configuration of selected extant taxa (typically, cephalochordates, gnathostomes, mammals and fish) as cladistic characters (Fig. 1). At least 500 million years ago (MYA), probably before the evolutionary split between jawed and jawless fishes, the single ancestral chordate Hox cluster—composed of all 13 *Hox* genes—underwent two rounds of genome duplication to give rise first to two proto clusters (AB and CD, Zhang and Nei, 1996), and then to the four clusters typically found in mammals (A, B, C, D) (Amores *et al.*, 1998; Holland and Garcia-Fernandez, 1996; Meyer, 1998; Meyer and Malaga-Trillo, 1999). The combined pattern of A cluster genes present in all vertebrates studied to date strongly suggests that the ancient gnathostome A cluster suffered a reduction that deleted paralogous groups 8 and 12 early in the evolution of vertebrates. Based on the shared absence of group 12 genes in all known A and B clusters, it must be assumed that this gene was lost in the proto AB cluster, after the first round of genome duplication. The loss of the group 8 gene must have therefore occurred after the second genome duplication but before the evolution of cartilaginous fishes around 460 MYA, since the absence of group 8 in the horn shark A cluster homologue is confirmed by complete sequencing of this genomic region (Kim *et al.*, 2000). Because the shark, lungfish and mouse A clusters do not show additional losses relative to the hypothetical gnathostome ancestor, the configuration of the A cluster in sarcopterygian

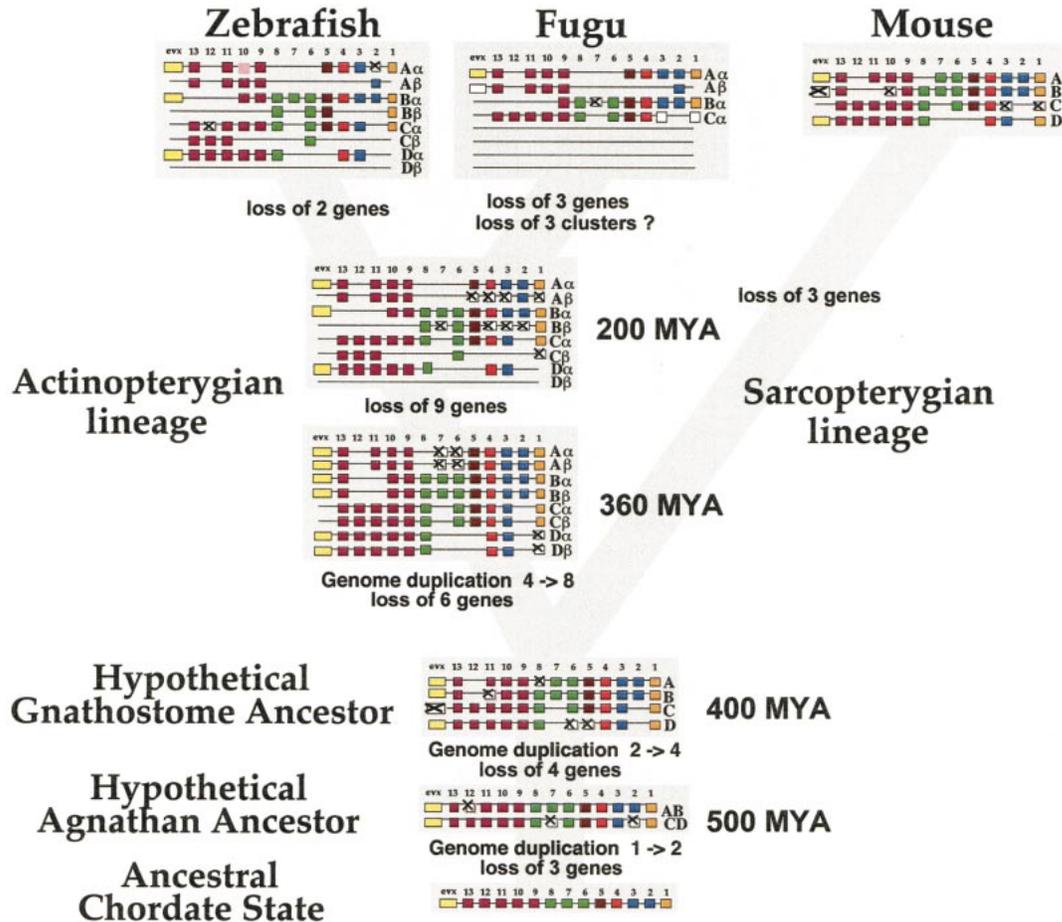


FIG. 1. A hypothetical scenario for the evolution of vertebrate Hox gene clusters, as inferred from the known Hox cluster architectures of *Amphioxus*, mouse, zebrafish and pufferfish (*Fugu*). The reconstruction was made using cladistic analysis, assuming *Amphioxus* as the ancestral chordate state and mapping Hox cluster evolution onto an expected vertebrate phylogeny. Colored boxes represent individual paralogous genes (1–13); boxes with crosses represent inferred gene losses. Clusters are labeled A–D, and  $\alpha$  or  $\beta$  are used to designate the duplicated clusters of fish. Approximate phylogenetic timing of the genome duplications and gene losses are indicated in million years ago (MYA).

and chondrichthyan lineages is likely to have remained unchanged since then. Extensive work on additional phylogenetically important lineages is desirable to be able to rule out that other independent gene losses occurred. Of particular interest is the configuration of basal sarcopterygian lineages such as the lungfish, for which only fragmentary data are available at present (Longhurst and Joss, 1999).

Another issue that needs further clarification is the finding of—so far—only two Hox clusters in the horn shark which show

strong affinities for the mammalian A and D clusters (Kim *et al.*, 2000). If the Hox complement of cartilaginous fish consisted of only two clusters, then these M (A-like) and N (D-like) clusters would be expected to correspond to the proto AB and CD clusters that probably originated after the first vertebrate genome duplication (Fig. 1). However, this is not likely to be the case, since the proto AB cluster is expected to contain a group 8 gene (absent in the shark M cluster, Fig. 1). In addition, the homology of the shark N cluster remains unclear:

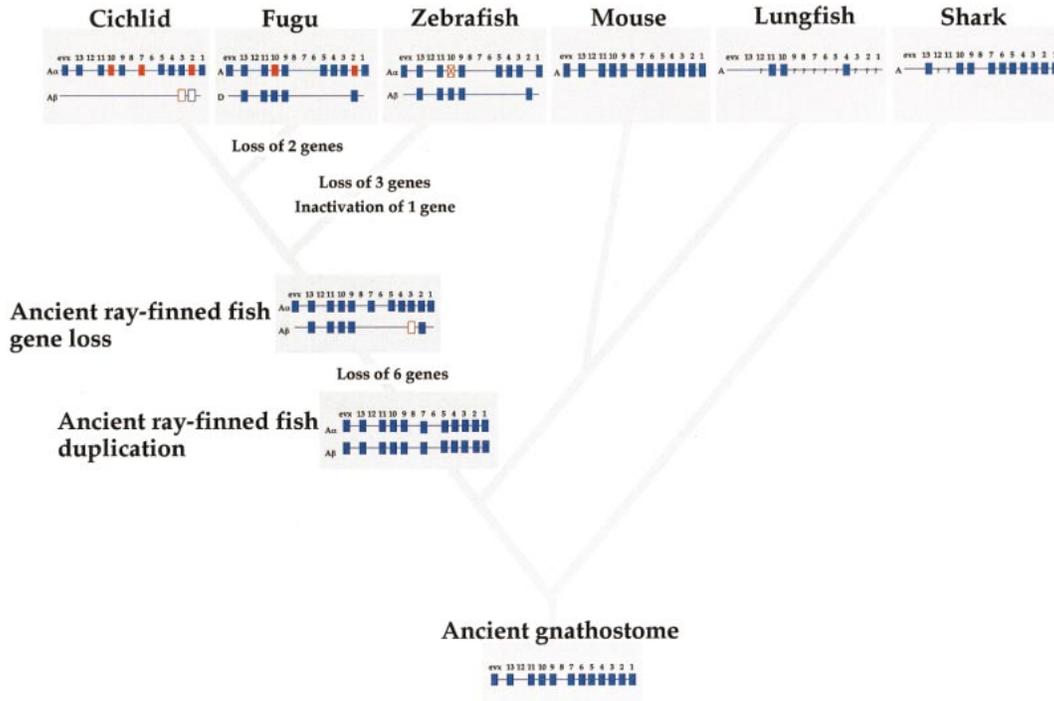


FIG. 2. Cladistic reconstruction of the evolution of Hox A clusters in vertebrates, based on the same taxa as in Figure 1, but with the addition of recent Hox gene data from the shark, lungfish, Medaka, striped bass and the cichlid fish *Oreochromis niloticus* (see text). The evolution of Hox A clusters was mapped onto an expected vertebrate phylogeny (independent of Hox genes). The independent gene losses that took place in zebrafish and *Fugu*—relative to cichlids—are indicated on the branches leading to these lineages. Solid boxes represent individual genes. Duplicated clusters are designated  $\alpha$  or  $\beta$ . *Fugu* clusters appear with their original names, A and D, but are now known to be the homologues of the A $\alpha$  and A $\beta$  clusters of zebrafish. Genes which have not been completely sequenced are indicated by open boxes. Genes where differential evolution between fish lineages has taken place are indicated with orange boxes; pseudogene *a10 $\alpha$*  in zebrafish is marked with a cross. Question marks represent non-characterized genomic regions.

phylogenetic analyses of this cluster's *Hox9* and *Evx* sequences as well as the absence of a group 6 gene support its orthology to D clusters (Kim *et al.*, 2000), but the presence of a group 5 gene does not. If one assumes that gene loss in the early gnathostome ancestor occurred rapidly after the two-to-four cluster duplication (and therefore before the evolution of the shark lineage), then the presence of a group 5 gene in the N cluster would be indicative of C-like affinities, as this paralogous group is found in all vertebrate C but not D clusters known to date. The true presence of a *hoxd5* gene in the shark would argue that this gene arose *de novo*. Since independent gene loss is much more likely to occur than independent gene gain, it would be more

parsimonious to assume that the shark N cluster is a C-like cluster which secondarily lost its group 6 gene (normally present in C but not in D clusters). The characterization of Hox clusters in the horn shark is likely to be incomplete, possibly due to the low coverage (2x) of the PAC (P1 artificial chromosome) genomic library utilized for these experiments, but we expect that additional work will eventually uncover the two remaining clusters.

In striking contrast to the conservation of Hox A cluster architecture observed in sarcopterygian and chondrichthyan genomes, actinopterygian Hox complexes show signs of a more eventful history that resulted in distinct cluster architectures for the different descending lineages (Fig. 2). The shared

presence of duplicated Hox A clusters in the fugu, zebrafish and cichlid lineages strongly supports the assertion that an additional genome duplication took place in the lineage leading to all modern bony fishes. The combined pattern of A genes present in the three distantly related fish orders strongly suggest that this ancestor possessed two A clusters ( $\alpha$  and  $\beta$ ) which lacked paralogous groups 8 and 12, like the ancient gnathostome ancestor, but also without group 6 copies. The shared loss of group 6 *Hox* genes in modern teleost  $A\alpha$  and  $A\beta$  clusters indicates that their common ancestor also lacked these two genes. However, it is presently difficult to establish the phylogenetic timing when this loss occurred. It is likely that both copies were lost in the common ancestor of the  $A\alpha$  and  $A\beta$  clusters, before the teleost genome duplication but after the evolution of tetrapods, since tetrapods do not have two A clusters. On the other hand, both copies could have been lost independently after the genome duplication but before the teleost radiation, which is a less parsimonious scenario that would constitute the only known case where two duplicated copies of a paralogous group were deleted simultaneously and independently. A third, perhaps more likely scenario is that the assumption of the absence of *a6* genes in the teleost ancestor is a reconstruction artifact. If so, a copy of the *a6* gene might still be present in other teleost lineages, which would imply that this copy was lost secondarily in the zebrafish, *Fugu* and cichlid lineages. In this respect, it is noteworthy to consider the case of the *a7* gene (Fig. 2). Based only on the *Fugu* and zebrafish architectures, it had to be assumed that both  $\alpha$  and  $\beta$  copies were independently (and possibly simultaneously) lost once in the teleost ancestor. However, identification of the *a7 $\alpha$*  gene in cichlid fish (Málaga-Trillo, Amores, McAndrew, Postlethwait and Meyer, unpublished data) and in the striped bass (Snell *et al.*, 1999), allows to ascertain the presence of this gene copy in the reconstructed configuration of the common teleost ancestor. Thus, the absence of both *a7* copies in *Fugu* zebrafish is minimized to the independent secondary loss of one paralogous

gene copy in both lineages. It is uncertain whether this shared loss of *a7 $\alpha$*  relative to the cichlid lineage is phylogenetically informative since current knowledge of fish phylogenies strongly suggests the relationships indicated in Figure 2. In addition to the possible shared loss of *Hox a6 $\alpha$*  and  $\beta$  genes, the Hox cluster of the teleost ancestor underwent deletions of paralogous groups 1, 3, 4, 5 and 7 in the  $A\beta$  cluster, leaving only groups 2, 9, 10, 11, and 13 as duplicated  $A\alpha$  and  $A\beta$  genes.

The data for the cichlid A clusters have not been previously published and will be presented elsewhere in more detail, as part of a more comprehensive description of cichlid Hox genes. However, some general features of the already fully characterized cichlid  $A\alpha$  cluster are relevant to this discussion. A set of overlapping cosmid clones covering this genomic region were subcloned and sequenced to completion. The entire DNA sequence obtained for the  $A\alpha$  cluster provides unambiguous information about cluster length, gene content and spacing, as well as non-coding regions. The cichlid  $A\alpha$  cluster contains paralogous groups 1, 2, 3, 4, 5, 7, 9, 10, 11, 13, and includes an *Evx* gene. All these genes have intact reading frames and the entire cluster extends over approximately 85 kilobases. This size is larger than that of the pufferfish (76 kb), but smaller than that of the shark (96 kb), and the relative spacing between genes ranges from 4 to 11 kilobases. The general pattern of gene content is typical for A clusters and no specific gene losses relative to other teleost fish are observed.

The most surprising observation when comparing the complete datasets for zebrafish, *Fugu* and cichlid Hox  $A\alpha$  clusters is that they vary at all. In addition to the extra genome duplication and gene loss that occurred in the teleost ancestor, its descendent lineages apparently experienced each their own particular evolutionary changes. The zebrafish lineage lost *a2 $\alpha$*  and *a7 $\alpha$* , and the pufferfish lost only *a7 $\alpha$* , but cichlid fish retained both genes. Also, *a10 $\alpha$*  appears to have undergone lineage specific changes. While this gene is intact in cichlids and the pufferfish, it turned into a pseudogene in the zebrafish. The reason for this differen-

tial evolution in paralogous groups 2, 7 and 10 of teleost A $\alpha$  clusters is not obvious, and detailed studies on the particular functions of these genes will be required to find out if their loss can be associated with selective consequences. The contrasting Hox cluster architectures of zebrafish, pufferfish and cichlid fish strongly suggest that independent (post-genome duplication) gene losses have taken place at different rates in these lineages since they arose as part of the teleost radiation, approximately 200 MYA (Pough *et al.*, 1999). The unexpected differences discussed here reveal an unprecedented degree of variability in Hox cluster gene composition, and suggest that even more variation can be uncovered in the genomes of other fish lineages. Unfortunately, most of the PCR studies carried out in other fishes are not extensive enough to detect these important differences in numbers of genes and clusters, and new laborious genomic surveys will be required to provide the awaited answers.

The strong contrast between the evolutionary histories of Hox clusters in actinopterygian and sarcopterygian lineages reveals two opposing but equally successful strategies: conservation of genomic structure *vs.* modification, variation and experimentation. The reasons why natural selection favored—or at least permitted—such divergent strategies in each lineage are not clear, but could be related to the different developmental and morphological challenges that each of these animal groups had to face in the course of evolution (Meyer and Schartl, 1999; Wittbrodt *et al.*, 1998).

#### HOX GENE SEQUENCE EVOLUTION IN FISH

An immediate consequence of genome duplications is genetic redundancy, the presence of more than one gene that perform the same function (see Nowak *et al.*, 1997; and Brookfield, 1997). Genetic redundancy is often regarded as beneficial because it increases an organism's genetic complexity by providing new genes that can potentially diverge in function (Ohno, 1970; Lynch and Conery, 2000). However, genetic redundancy through polyploidization may also result in dosage effects (Guo *et al.*, 1996), a consequence which, at least

for some genes, might lead to deleterious consequences associated with overexpression (Krebs and Feder, 1997; Zelinski *et al.*, 2001). Let us consider the hypothetical scenario of an organism that develops immediately after a genome duplication took place. The functional consequences of suddenly having two identical copies of every locus obviously depend on the specific function of each given gene product. The likelihood of keeping or eliminating newly duplicated gene copies will therefore depend on how the traits determined by these genes are affected by a gene duplication. Duplicated loci for which a quantitative increase of gene products would be beneficial for the individual, are likely to be maintained, whereas duplicated loci whose increased activity would result in deleterious effects, or even lethality, are expected to be selected against. In this latter case, the (sudden) requirement to avoid deleterious levels of gene expression would lead to selection for the elimination of one copy or for accelerated changes in both copies so that they subdivide the role of their single pre-duplication ancestor (“subfunctionalization,” see Force *et al.*, 1999). In this regard, the loss/inactivation of different *Hox* gene duplicates in different fish lineages could be indicative of selective pressure against the preservation of identical gene copies at those particular loci (although the traditional explanation for duplicate gene loss would be that additional copies are eliminated through random drift because their inactivation has little or no effect on the phenotype).

During evolution, Actinopterygian Hox clusters have clearly experienced gene losses (after the initial doubling of their number) at a much higher rate than Sarcopterygian lineages, where no variation in cluster architecture has been reported. This difference in the rate of gene loss possibly reflects a general adaptive tendency towards rapid evolution after genome duplication (*e.g.*, increased rates of gene deletion). It is known that genome duplications can cause genomic instability (Matzke *et al.*, 1999), which in turn can lead to rapid structural changes. Such changes may involve the programmed loss of subgenomic sequences

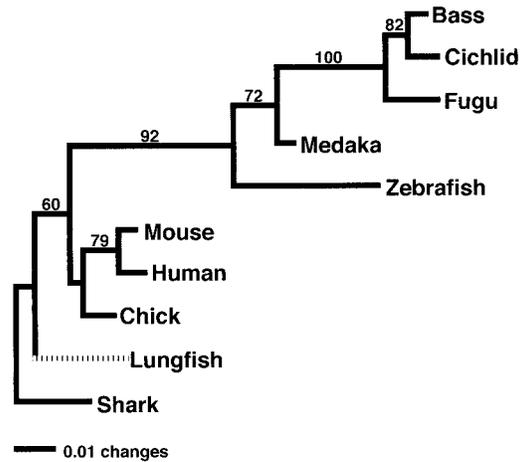


FIG. 3. Neighbor joining phylogenetic tree of vertebrate *hoxa4/a9* homeodomain aminoacid sequences. Numbers above the branches indicate bootstrap support values (1,000 replications). The length of the branches is proportional to the genetic distance between taxa (mean character difference) and reflect the differences in rates between sarcopterygian and actinopterygian lineages, since they are sister lineages.

in order to allow diploid meiotic behavior, thereby generating extensive genetic diversity (Song *et al.*, 1995).

In the absence of comparative genomic and developmental studies, it is not yet possible to establish at which rate the functions of Actinopterygian *Hox* genes might have been modified or entirely replaced by new ones. However, it is possible to analyze their rates of gene sequence evolution and determine whether rapid changes have occurred in a particular lineage. We examined and compared the variation in *Hox* gene sequences among the vertebrate groups discussed in this study. The neighbor joining tree in Figure 3 presents a vertebrate phylogeny based on the alignment of *hoxa4* and *hoxa9* homeodomain aminoacid sequences, using the horn shark as an outgroup. Only the position of the Australian lungfish, *Neoceratodus forsteri*, could not be accurately resolved because of the insufficient sequence data available for this taxon (*hoxa9* gene not reported yet). The rest of the taxa are grouped into two distinct lineages, containing sarcopterygian (chick, mouse and human) and actinopterygian sequences (teleost fish), respectively. The re-

lationships among teleost lineages in our molecular phylogeny are supported with high bootstrap values and agree with traditional expectations derived from morphological characters (Nelson, 1994). Perciformes (cichlid and bass) and Tetraodontiformes (pufferfish) are more closely related to each other than to Cyprinodontiformes (Medaka) and to the more distantly related Cypriniformes (zebrafish). Closer examination of the actinopterygian and sarcopterygian sister groups reveals another striking difference between the two (equally old) groups, the former having longer branches (genetic distances) than the latter. This result indicates that after the split of the two lineages, actinopterygian *Hox* sequences accumulated differences at a much higher rate than those of their sarcopterygian homologues, *i.e.*, the *hoxa4/hoxa9* gene sequences do not seem to have evolved at constant rates in all vertebrates. This is also supported by a Kishino-Hasegawa (1996) test as implemented in the Puzzle program (Strimmer and von Haeseler, 1996). Maximum likelihood estimates of branch lengths based on the Dayhoff substitution model (Dayhoff, 1978) and assuming site rate homogeneity clearly rejects a clock-like tree at a significance level of 5%.

The disparity of evolutionary rates suggests that the *Hox* genes of these two vertebrate lineages might have been subjected to different selective pressure. Interestingly, the rapid evolution of teleost *Hox* sequences coincides with the occurrence of an additional genome duplication and a subsequent higher rate of gene loss in this group. Although the precise timing of all of these events is not clear, it is tempting to suggest that this link might be causal. Positive selection for gene loss and increased levels of DNA sequence divergence could have effectively reduced the levels of genetic redundancy created by a genome duplication. Thus, the generation of genomic diversity in the *Hox* clusters of teleost fish would be just one example of how organisms independently face the consequences of genomic instability after genome duplication. The study of the possible effects of these genetic changes on the evolution of morphological diversity, by performing detailed

comparative analyses of gene expression and function in selected teleost fish, will permit further testing on this.

## ACKNOWLEDGMENTS

The sequence data on cichlid Hox gene clusters from this paper are not yet published. We gratefully acknowledge John Postlethwait and Angel Amores at the University of Oregon for their involvement in this project, as well as our collaborators at the DOE Joint Genome Institute, for contributing large amounts of DNA sequence data. We also would like to thank Yves van de Peer for his input on evolutionary rates, John Taylor for critical discussion of the manuscript, as well as the Fond der Chemischen Industrie, Deutsche Forschungsgemeinschaft (DFG) and the University of Konstanz for financial support.

## REFERENCES

- Akam, M. 1989. Hox and HOM: Homologous gene clusters in insects and vertebrates. *Cell* 57:347–349.
- Amores, A., A. Force, Y. L. Yan, L. Joly, C. Amemiya, A. Fritz, R. K. Ho, J. Langeland, V. Prince, Y. L. Wang, M. Westerfield, M. Ekker, and J. H. Postlethwait. 1998. Zebrafish hox clusters and vertebrate genome evolution. *Science* 282:1711–1714.
- Aparicio, S. 2000. Vertebrate evolution: Recent perspectives from fish. *Trends Genet.* 16:54–56.
- Aparicio, S., K. Hawker, A. Cottage, Y. Mikawa, L. Zuo, B. Venkatesh, E. Chen, R. Krumlauf, and S. Brenner. 1997. Organization of the Fugu rubripes Hox clusters: Evidence for continuing evolution of vertebrate Hox complexes. *Nat. Genet.* 16:79–83.
- Brookfield, J. F. Y. 1997. Genetic redundancy. *Adv. Genet.* 36:137–155.
- Dayhoff, M. O. 1978. *Protein segment dictionary 78: From the Atlas of protein sequence and structure, volume 5, and supplements 1, 2, and 3.* National Biomedical Research Foundation; Georgetown University Medical Center, Silver Springs Md. Washington, D.C.
- De Robertis, E. M. 1997. Evolutionary biology. The ancestry of segmentation. *Nature* 387:25–26.
- Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan, and J. Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545.
- Gehring, W. J. 1998. *Master control genes in development and evolution: The homeobox story.* Yale University Press, New Haven.
- Guo, M., D. Davis, and J. A. Birchler. 1996. Dosage effects on gene expression in a maize ploidy series. *Genetics* 142:1349–1355.
- Holland, P. W. 1997. Vertebrate evolution: Something fishy about Hox genes. *Curr. Biol.* 7:R570–R572.
- Holland, P. W. and J. Garcia-Fernandez. 1996. Hox genes and chordate evolution. *Dev. Biol.* 173:382–395.
- Holland, P. W., J. Garcia-Fernandez, N. A. Williams, and A. Sidow. 1994. Gene duplications and the origins of vertebrate development. *Dev. Suppl.* 125–133.
- Kim, C. B., C. Amemiya, W. Bailey, K. Kawasaki, J. Mezey, W. Miller, S. Minoshima, N. Shimizu, G. Wagner, and F. Ruddle. 2000. Hox cluster genomics in the horn shark, *heterodontus francisci*. *Proc. Natl. Acad. Sci. U.S.A.* 97:1655–1660.
- Kishino, H. and M. Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J. Mol. Evol.* 29:170–179.
- Krebs, R. A. and M. E. Feder. 1997. Deleterious consequences of Hsp70 overexpression in *Drosophila melanogaster* larvae. *Cell Stress Chaperones* 2(1):60–71.
- Krumlauf, R. 1994. Hox genes in vertebrate development. *Cell* 78:191–201.
- Kurosawa, G., K. Yamada, H. Ishiguro, and H. Hori. 1999. Hox gene complexity in medaka fish may be similar to that in pufferfish rather than zebrafish. *Biochem. Biophys. Res. Commun.* 260:66–70.
- Lewis, E. B. 1978. A gene complex controlling segmentation in *Drosophila*. *Nature* 276:565–570.
- Longhurst, T. J. and J. M. Joss. 1999. Homeobox genes in the Australian lungfish, *Neoceratodus forsteri*. *J. Exp. Zool.* 285:140–145.
- Lynch, M. and J. S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1154.
- Matzke, M. A., O. M. Scheid, and A. J. M. Matzke. 1999. Rapid structural and epigenetic changes in polyploid and aneuploid genomes. *Bioessays* 21:761–767.
- McGinnis, W. and R. Krumlauf. 1992. Homeobox genes and axial patterning. *Cell* 68:283–302.
- Meyer, A. 1996. The evolution of body plans: HOM/Hox cluster evolution, model systems and the importance of phylogeny. In P. H. Harvey, A. J. Leigh Brown, J. M. Smith, and S. Nee (eds.), *New uses for new phylogenies*, pp. 322–340. Oxford University Press.
- Meyer, A. 1998. Hox gene variation and evolution. *Nature* 391:225, 227–228.
- Meyer, A. and E. Malaga-Trillo. 1999. Vertebrate genomics: More fishy tales about Hox genes. *Curr. Biol.* 9:R210–R213.
- Meyer, A. and M. Schartl. 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, B. Y., M. J. Blanco, and G. P. Wagner. 1996. PCR-survey of Hox-genes of the zebrafish: New sequence information and evolutionary implications. *J. Exp. Zool.* 274:193–206.
- Misof, B. Y. and G. P. Wagner. 1996. Evidence for four

- Hox clusters in the killifish *Fundulus heteroclitus* (teleostei). *Mol. Phylogenet. Evol.* 5:309–322.
- Naruse, K., S. Fukamachi, H. Mitani, M. Kondo, T. Matsuoka, S. Kondo, N. Hanamura, Y. Morita, K. Hasegawa, R. Nishigaki, A. Shimada, H. Wada, T. Kusakabe, N. Suzuki, M. Kinoshita, A. Kanamori, T. Terado, H. Kimura, M. Nonaka, and A. Shima. 2000. A Detailed Linkage Map of Medaka, *Oryzias latipes*. Comparative genomics and genome evolution. *Genetics* 154:1773–1784.
- Nelson, J. S. 1994. *Fishes of the world*. 3rd ed. J. Wiley, New York.
- Nowak, M. A., M. C. Boerlijst, J. Cooke, and J. M. Smith. 1997. Evolution of genetic redundancy. *Nature* 388:167–171.
- Ohno, S. 1970. *Evolution by gene duplication*. Springer-Verlag Berlin, New York.
- Pavell, A. M. and E. J. Stellwag. 1994. Survey of Hox-like genes in the teleost *Morone saxatilis*: Implications for evolution of the Hox gene family. *Mol. Mar. Biol. Biotechnol.* 3:149–157.
- Pough, F. H., C. M. Janis, and J. B. Heiser. 1999. *Vertebrate life*. 5th ed. Prentice Hall, Upper Saddle River, N.J.
- Prince, V. E., L. Joly, M. Ekker, and R. K. Ho. 1998. Zebrafish hox genes: Genomic organization and modified colinear expression patterns in the trunk. *Development* 125:407–420.
- Ruddle, F. H., K. L. Bentley, M. T. Murtha, and N. Risch. 1994. Gene loss and gain in the evolution of the vertebrates. *Dev. Suppl.* 155–161.
- Slack, J. M., P. W. Holland, and C. F. Graham. 1993. The zootype and the phylotypic stage. *Nature* 361:490–492.
- Snell, E. A., J. L. Scemama, and E. J. Stellwag. 1999. Genomic organization of the Hoxa4-Hoxa10 region from *Morone saxatilis*: Implications for Hox gene evolution among vertebrates. *J. Exp. Zool.* 285:41–49.
- Song, K., P. Lu, K. Tang, and T. C. Osborn. 1995. Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proc. Natl. Acad. Sci. U.S.A.* 92:7719–7723.
- Stellwag, E. J. 1999. Hox gene duplication in fish. *Semin. Cell Dev. Biol.* 10:531–540.
- Strimmer, K. and A. von Haeseler. 1996. Quartet puzzling: A quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* 13:964–969.
- Wittbrodt, J., A. Meyer, and M. Scharl. 1998. More genes in fish? *Bioessays* 20:511–515.
- Zelinski, D. P., N. D. Zantek, J. C. Stewart, A. R. Izarray, and M. S. Kinch. 2001. EphA2 overexpression causes tumorigenesis of mammary epithelial cells. *Cancer Res.* 61(5):2301–2306.
- Zhang, J. and M. Nei. 1996. Evolution of Antennapedia-class homeobox genes. *Genetics* 142:295–303.