First publ. in: Current Biology 9 (1999), 6, pp. R210-R213

Vertebrate genomics: **More fishy tales about *Hox* genes**

Axel Meyer and Edward Málaga-Trillo

Zebrafish *Hox* genes are arranged in at least seven clusters, rather than the four clusters typical of vertebrates. This suggests that an additional genome duplication occurred on the fish lineage and explains why many gene families are typically about half the size in land vertebrates than they are in fish.

Address: Department of Biology, University of Konstanz, 78457 Konstanz, Germany.
E-mail: axel.meyer@uni-konstanz.de

Current Biology 1999, 9:R210–R213
http://biomednet.com/elecref/09609822009R0210

© Elsevier Science Ltd ISSN 0960-9822

Comparative developmental biology is experiencing something of a renaissance. One reason for this is the hope that our understanding of the relationship between genotype and phenotype will be improved by studying patterns of diversification, the origin of morphological novelties and the evolution of body plans. Particular hope has been pinned on the *Hox* genes, which appear to play pivotal roles in specifying the body plans of a wide range of metazoan species [1]. The *Hox* genes, which encode transcription factors, are arranged in genomic clusters that are strikingly colinear with their spatial and temporal expression patterns. Thus, at one end of a cluster are located the 'anterior' *Hox* genes, which are expressed earlier in development and more anteriorly along the main body axis, whereas at the other end of a cluster are the 'posterior' genes, which are switched on later in development and in more distal portions of the body.

It has been suggested that increasing complexity of body plans during evolution might be causally correlated with increasing complexity of the *Hox* complexes (see [2] for example). Invertebrates have only a single *Hox* gene complex, and the common ancestor of all chordates probably also had only a single cluster [3]. During the evolution of chordates from relatively simple cephalochordates such as Amphioxus (Figure 1) to more complex organisms such as mammals, the single ancestral *Hox* cluster was probably duplicated twice, giving the four *Hox* clusters (clusters A–D) seen in the human and mouse genomes. These duplications from one to two, and then from two to four, clusters probably occurred, not as tandem gene duplications but either as a result of individual chromosome duplications or, more likely, whole-genome duplications, as the clusters are each on different chromosomes. Mice and men both have 39 *Hox* genes in four clusters, and this was expected to be the typical *Hox* repertoire and arrangement for all vertebrates. But as we shall see, this is proving not to be the case.

It turns out that neither the number of *Hox* genes nor the number of *Hox* clusters is fixed among chordates. This was first shown by Aparicio *et al.* [4], who found that the puffer fish *Fugu* has only 31 *Hox* genes — rather than the expected 39, typical of land vertebrates — arranged in four gene clusters. As *Fugu* has not only an unusually small genome, but also an almost aberrantly reduced morphology — it lacks several sets of fins and bones — the concomitant reduction in the number of *Hox* genes seemed to make some sense. It is, however, quite unclear whether the loss of *Hox* genes is in any way related to the secondary simplification and loss of some morphological structures [4–6]. And now the cosy view that vertebrates — both land vertebrates and fishes — all have four homologous *Hox* clusters [4,7] has also been challenged [8,9].

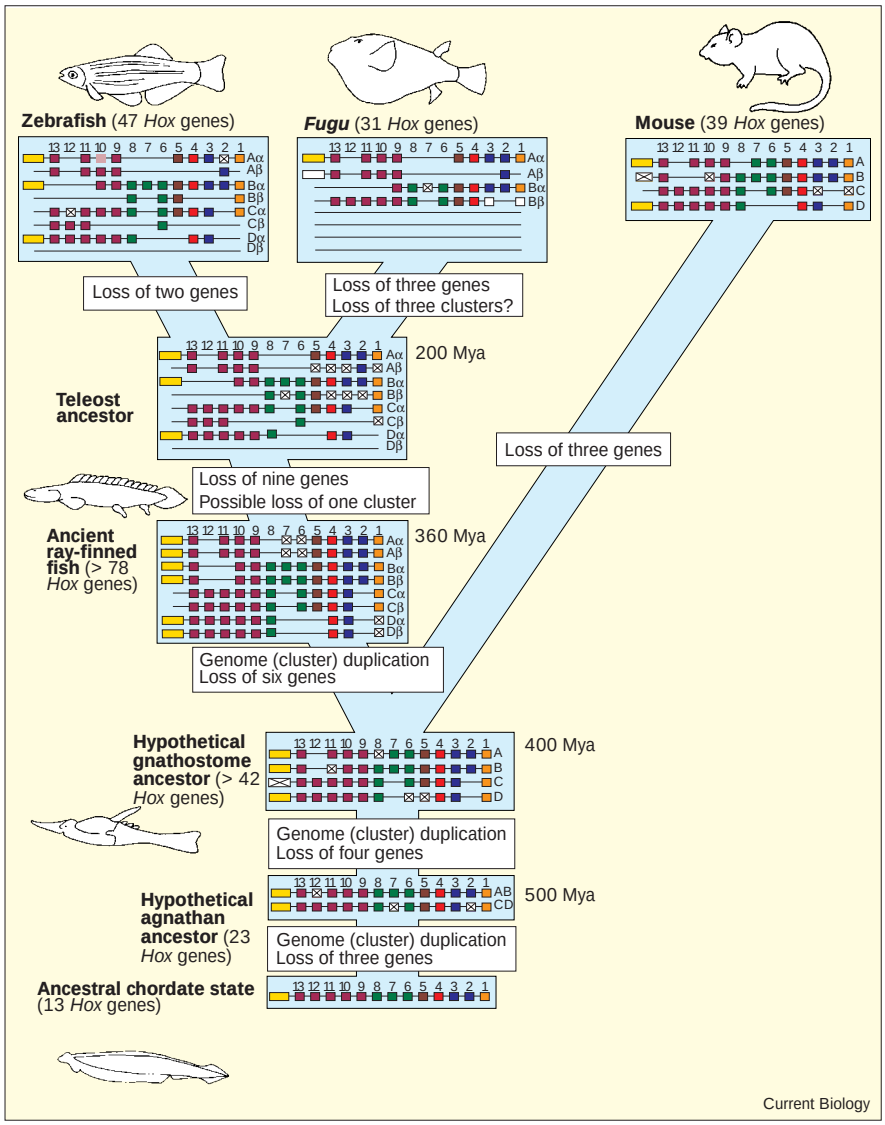
Prince *et al.* [8] recently described 42 *Hox* genes from the zebrafish and suggested that they are arranged in six, rather than the expected four, clusters. Apparently, in the evolutionary lineage leading to zebrafish, some *Hox* gene clusters must have been duplicated to bring the total number to at least six, two more than had previously been found in any other vertebrate. Amores *et al.* [9] have now clarified this puzzling situation by their discovery of at least 48 *Hox* genes in the zebrafish, clustered in at least seven *Hox* gene complexes — three more than the typical vertebrate number of four. In this study, zebrafish genomic DNA was cloned in P1 artificial chromosomes (PACs) and *Hox* genes were amplified using the polymerase chain reaction (PCR) with degenerate primers. The amplified genes were then mapped to zebrafish chromosomes. The results may demand a reevaluation of ideas about the evolutionary link between complexity of body plans and complexity of genomic organization, and also provide a testable model for the evolutionary history of the chordate genome.

From their observations on the genomic organization of the zebrafish *Hox* gene clusters, Amores *et al.* [9] have developed a model for the evolution of the *Hox* complexes in vertebrates (outlined in Figure 1). This model is based on the shared presence of certain genes, and the assumption that genes can be lost independently but are much less likely to arise independently. The single ancestral *Hox* cluster architecture of the common ancestor of all chordates is composed of the full complement of 13 *Hox* genes, as well as genes such as *even-skipped* (*Evx*) that are linked to this *Hox* complex. This is

Konstanzer Online-Publikations-System (KOPS)
URL: http://www.ub.uni-konstanz.de/kops/volltexte/2007/3532/
URN: http://nbn-resolving.de/urn:nbn:de:bsz:352-opus-35327

Figure 1

A possible scheme for the evolution of the *Hox* gene clusters during vertebrate evolution that takes into account recent data on *Hox* gene organization in fish [4,7–9]. Rectangles with crosses represent inferred gene losses. In the *Fugu* genome, *HoxCα1* and *HoxCα3* can be recognized but are only pseudogenes (white squares). The phylogenetic timing of cluster duplications and gene losses is indicated on the tree. Mya, million years ago.

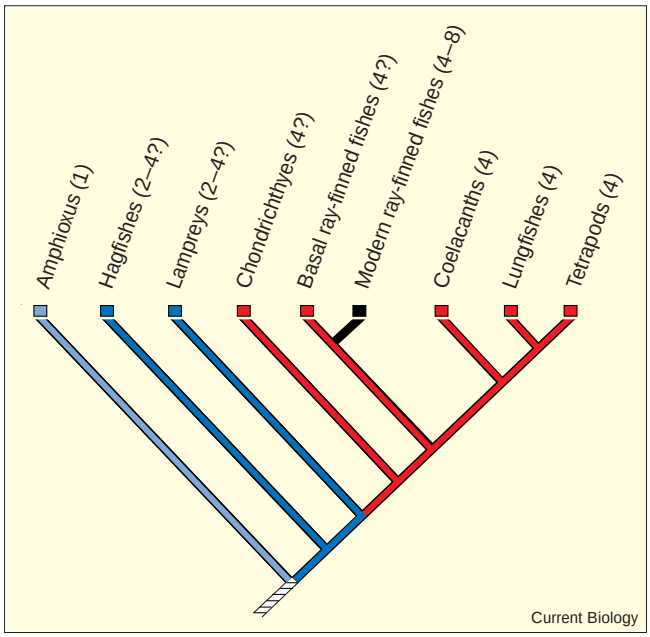


the genomic architecture found in the living cephalochordate Amphioxus [3] (Figure 1).

According to the scenario described by Amores *et al.* [9], early in the evolution of vertebrates — at least 500 million years ago, before the evolution of jawless (agnathan) fishes — this single *Hox* cluster duplicated to form two clusters, AB and CD. Later in vertebrate evolution, but perhaps still before the last common ancestor of all jawed vertebrates arose more than 400 million years ago, these two clusters duplicated again, most likely through an entire genome duplication, to give the four clusters A, B, C and D (Figures 1,2). The A and B clusters are derived from the ancestral AB cluster, and the C and D clusters from the hypothetical CD cluster. Such an evolutionary relationship among the four *Hox* clusters had been suggested before ([10] for example), but it could not be firmly established prior to the new work on the zebrafish *Hox* clusters [9], because of the lack of sufficient phylogenetically informative data [2,11–13].

Because the zebrafish seems to have at least seven, perhaps eight (A. Amores, personal communication), *Hox* clusters rather than the expected four, it is likely that an additional genome duplication took place along the lineage leading to teleost fishes (Figure 1). This third genome duplication apparently occurred less than 360 million years ago, after the lineage leading to teleost fishes separated from that leading to coelacanths, lungfishes and land vertebrates [14] (Figures 1,2). Species on our evolutionary branch, after the split from the common ancestor of the ray-finned fishes, are all likely to have

Figure 2



Phylogeny of the major chordate groups, with estimated numbers of *Hox* gene clusters indicated in brackets. Only a few of these numbers are known with certainty from genomic analyses – Amphioxus has one, some tetrapods have four and some modern ray-finned fishes have at least four, zebrafish having seven. The other numbers are estimated from partial knowledge of the *Hox* cluster numbers and a most-parsimonious reconstruction of the major evolutionary events in chordate genome evolution based on the illustrated phylogenetic topology. Lineages with the same colors are likely to have the same numbers of *Hox* gene clusters. Changes in the colors of branches – light blue to dark blue, dark blue to red, and red to black – indicate the likely phylogenetic timing of genome duplication events. The base of this tree is stippled, as it is not yet known whether the common ancestor of chordates already had a chordate *Hox* cluster architecture as present in Amphioxus.

only four *Hox* complexes (Figure 2), except for certain polyploid tetrapods.

It should be possible to test the veracity of this model, as well as the phylogenetic timing of the *Hox* cluster duplications, by comparative genomic analyses in the appropriate species. This will require, however, accurate knowledge of the phylogenetic relationships among fish lineages [11]. Several important evolutionary lineages that are not represented in Figure 1 may end up holding the key to testing this model (Figure 2). The hagfishes (Myxiniformes) and lampreys (Petromyzontiformes), which diverged early on from the main lineage leading to modern fish, are likely to prove particularly informative. There is still some controversy over whether hagfishes and lampreys form a monophyletic group [15], so that they are likely to have the same number of *Hox* clusters, or do not (as illustrated in Figure 2), in which case it is possible at that the hagfishes have two or three clusters and the lampreys four.

The most basal of the living jawed fish, the cartilaginous fish (Chondrichthyes), branched off from the lineage leading to both modern bony fishes and the tetrapods before the most basal bony fish, the bichirs (Polypteriformes) and sturgeons (Acipenseriformes), branched off from the main lineage about 400 million years ago. These basal jawed fish might have a *Hox* cluster arrangement similar to the hypothetical one suggested in Figure 1 for an 'ancient ray-finned fish'. It should be noted that the phylogenetic position of bichirs is still uncertain; even a phylogenetic analysis based on an entire mitochondrial genome could not unequivocally place the bichirs with either the ray-finned fish or lobe-finned fish [16], yet their phylogenetic position is crucial if we wish to understand the evolution of vertebrate structures, such as limbs, and genomes.

More derived fish than these, the bow fins and also the gars (Lepiseusteiformes), might have a genomic architecture similar to that found in all modern fishes, as illustrated for the 'teleost ancestor' in Figure 1. Lungfish, probably our closest living relatives among the fish [14], and the coelacanth (*Latimeria*) most likely will be found to have four clusters, as they branched off from the lineage leading to the tetrapods about 360 million years ago (Figure 2) [14,17]. It is possible, though it does not seem likely, that four clusters have been lost since the first vertebrate crawled on land, in which case the lungfishes and coelacanths may be found to have more than four clusters, possibly eight.

The new data [9] suggest that *Fugu* might have more *Hox* genes than have been found before, and that what was first called the D cluster in *Fugu* is actually the Aα cluster (A. Amores, personal communication). If that were so, it would be likely that an orthologous D cluster will still be found in *Fugu* (Figure 1). The high degree of variability among the *Hox* clusters even of such relatively closely related species such as zebrafish and *Fugu*, the common ancestor of which existed more recently than 200 million years ago, suggests that independent losses of *Hox* genes, and even entire *Hox* clusters, occurred in different lineages of fishes. This would raise the question of whether the patterns of *Hox* gene losses in different fish lineages tell us something about their function? It would seem oversimplistic, though not inconceivable, that losses of appendages — pelvic and pectoral fins have been lost repeatedly and independently in unrelated groups of fish, and the same is true of limbs in amniotes such as snakes and whales — are associated with different regulation or even losses of *Hox* genes. If this simple view is correct, we might be able to predict which *Hox* genes have been lost in which groups of fish on the basis of their phenotype.

From the available data, it would appear that losses of *Hox* genes and clusters have occurred at quite variable rates in different fish lineages. The zebrafish is likely to have lost

only about two *Hox* genes since it diverged from its common ancestor with *Fugu*. Losses seem to have occurred at a significantly faster rate on the *Fugu* lineage, which apparently lost three entire clusters as well as three genes on the remaining clusters (Figure 1). Remarkably, the *Hox* clusters of the mouse might be differentiated from those of its last common ancestor with the pufferfish and the zebrafish by only three *Hox* gene losses, even though this last common ancestor lived about 360–400 million years ago (Figure 1). It was recently suggested [18] that the apparent acceleration of genomic evolution and greater gene copy numbers in fish might be the result of a fish-specific genome duplication and be correlated with accelerated morphological evolution and speciation in fish [18]. Fish are the most successful group of vertebrates, comprising more than 20,000 species with vastly different morphologies, some of which could have arisen within very short evolutionary time spans [19].

How do genomes evolve? The clustering of *Hox* genes might be conserved because of the requirement to keep *Hox* genes reasonably close to shared regulatory elements. The zebrafish data show, however, that cluster number can vary quite dramatically, and that losses of individual *Hox* genes within clusters are common. That genes and even pseudogenes are apparently maintained in genomes for long evolutionary time spans might suggest that there is less of a genomic cost to redundancy and to maintaining, and possibly 'recycling', genetic programs for different purposes [20] than one might have expected [21]. This propensity of genes to stick around in genomes might increase the evolutionary potential of evolutionary lineages, such as the fish lineage, to respond more quickly and flexibly to changing ecological situations [18].

Acknowledgements

We would like to thank Heike Haunstetter for the drawing of the figures and Angel Amores and John Postlewait for discussions and permission to cite unpublished work. We also thank the US National Science Foundation, the Deutsche Forschungsgemeinschaft and the University of Konstanz for financial support.

References

1. Lewis E: A gene complex controlling segmentation in ***Drosophilia***. *Nature* 1978, 276:565-570.
2. Holland PWH, Garcia-Fernández J: ***Hox*** genes and chordate evolution. *Dev Biol* 1996, 173:382-395.
3. Garcia-Fernández J, Holland PWH: Archetypal organization of the amphioxus ***Hox*** gene cluster. *Nature* 1994, 370:563-566.
4. Aparicio S, Hawker K, Cottage A, Mikawa Y, Zuo L, Venketesh B, Chen E, Krumlauf R, Brenner S: Organization of the ***Fugu rubripes Hox*** clusters: evidence for continuing evolution of vertebrate ***Hox*** complexes. *Nat Genet* 1997, 16:79-83.
5. Holland PWH: Something fishy about ***Hox*** genes. *Curr Biol* 1997, 7:R570-R572.
6. Meyer A: ***Hox*** gene variation and evolution. *Nature* 1998, 391:225-228.
7. Misof BY, Blanco MJ, Wagner GP: PCR-survey of ***Hox***-genes of the zebrafish: new sequence information and evolutionary implications. *J Exp Zool* 1996, 274:193-206.
8. Prince VE, Jolly J, Ekker M, Ho R: Zebrafish ***hox*** genes: genomic organization and modified colinear expression patterns in the trunk. *Development* 1998, 125:407-420.
9. Amores A, Force A, Yan, Y-L, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang Y-L, *et al.*: Zebrafish ***hox*** clusters and vertebrate genome evolution. *Science* 1998, 282:1711-1714.
10. Kappen C, Ruddle FH: Evolution of a regulatory gene family: ***HOM/HOX*** genes. *Curr Opin Genet Dev* 1993, 3:931-938.
11. Meyer A: The evolution of body plans: ***HOM/Hox*** cluster evolution, model systems and the importance of phylogeny. In *New Uses for New Phylogenies.* Edited by Harvey PH, Leigh Brown AJ, Maynard Smith J, Nee S. Oxford: Oxford University Press; 1996:322-340.
12. Zhang J, Nei M: Evolution of ***antennapedia***-class homeobox genes. *Genetics* 1996, 142:295-303.
13. Bailey WJ, Kim, J, Wagner GP, Ruddle FH: Phylogenetic reconstruction of vertebrate Hox cluster duplications. *Mol Biol Evol* 1998, 14:843-853.
14. Zardoya R, Meyer A: Molecular phylogenetic information on the identity of the closest living relatives(s) of land vertebrates. *Naturwissenschaften* 1997, 84:389-397.
15. Mallat J, Sullivan J: 28S and 18S rDNA sequences support the monophyly of lampreys and hagfishes. *Mol Biol Evol* 1998, 15:1706-1718.
16. Noack K, Zardoya R, Meyer A: The complete mitochondrial DNA sequence of the bichir (***Polypterus ornatipinnis***), a basal ray-finned fish: ancient establishment of the consensus vertebrate gene order. *Genetics* 1996, 144:1165-1180.
17. Meyer A, Wilson AC: Origin of tetrapods inferred from their mitochondrial DNA affiliation to lungfish. *J Mol Evol* 1990, 31:359-364.
18. Wittbrodt J, Meyer A, Schartl M: More genes in fish? *BioEssays* 1998, 20:511-515.
19. Meyer A, Kocher TD, Basasibwaki P, Wilson AC: Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature* 1990, 347:550-553.
20. Meyer A: Homology and homoplasy: retention of ancient genetic programmes? In *Homology.* Edited by Bock G, Cardew G. London: Wiley; 1999:72-82.
21. Nowak MA, Boerlijst MC, Cooke J, Maynard Smith J: Evolution of genetic redundancy. *Nature* 1997, 388:167-171.