Evidence for Independent *Hox* gene duplications in the hagfish lineage: A PCR-based gene inventory of *Eptatretus stoutii*

Peter F. Stadler [a,d], Claudia Fried [a], Sonja J. Prohaska [a], Wendy J. Bailey [b], Bernhard Y. Misof [c], Frank H. Ruddle [e], Günter P. Wagner [f],

[a] *Lehrstuhl für Bioinformatik, Institut für Informatik, Universität Leipzig, Kreuzstraße 7b, D-04103 Leipzig, Germany.*
`{studla,claudia,sonja}@bioinf.uni-leipzig.de`

[b] *Dept. Molecular Profiling, Merck & Co., Inc. West Point, PA 19486, USA*
`wendy_bailey@merck.com`

[c] *Zoologisches Forschungsinstitut und Museum Alexander Koenig Sektion Niedere Arthropoden Adenauerallee 160, D-53113 Bonn, Germany*
`b.misof.zfmk@uni-bonn.de`

[d] *Institut für Theoretische Chemie und Molekulare Strukturbiologie, Universität Wien, Währingerstraße 17, A-1090 Wien, Austria*

[e] *Department of Molecular, Cellular and Developmental Biology Yale University, New Haven, CT 06520-8103, USA*
`frank.ruddle@yale.edu`

[f] *Department of Ecology and Evolutionary Biology Yale University, New Haven, CT 06520-8106, USA*
`gunter.wagner@yale.edu`

Abstract

Hox genes code for transcription factors that play a major role in the development of all animal phyla. In invertebrates these genes usually occur as tightly linked cluster, with a few exceptions where the clusters have been dissolved. Only in vertebrates multiple clusters have been demonstrated which arose by duplication from a single ancestral cluster. This history of Hox cluster duplications, in particular during the early elaboration of the vertebrate body plan, is still poorly understood. In this paper we report the results of a PCR survey on genomic DNA of the pacific hagfish *Eptatretus stoutii*. Hagfishes are one of two clades of recent jawless fishes that are an offshoot of the early radiation of jawless vertebrates. Our data provides evidence for at least 33 distinct Hox genes in the hagfish genome, which is most compatible

with the hypothesis of multiple Hox clusters. The largest number, seven, of distinct homeobox fragments could be assigned to paralog group 9, which could imply that the hagfish has more than four clusters. Quartet mapping reveals that within each paralog group the hagfish sequences are statistically more closely related to gnathostome Hox genes than with either amphioxus or lamprey genes. These results support two assumptions about the history of Hox genes: (1) The association of hagfish homeobox sequences with gnathostome sequences suggests that at least one Hox cluster duplication event happened in the stem of vertebrates, i.e., prior to the most recent common ancestor of jawed and jawless vertebrates. (2) The high number of paralog group 9 sequences in hagfish and the phylogenetic position of hagfish suggests that the hagfish lineage underwent additional independent Hox cluster/-gene duplication events.

Key words: Eptatretus stoutii, Hox genes, PCR survey, vertebrate evolution.

1 Introduction

Since Ohno's book on the role of gene duplication in evolution, the idea that gene and genome duplication played a major role in the origin of vertebrates has grown in support (Ohno, 1970). It is now clear that in fact vertebrates tend to have more copies of genes that have homologs in invertebrates and that there is also extensive variation in gene number among different clades of vertebrates (Meyer & Schartl, 1999). Furthermore, the age distribution of duplicated genes is consistent with at least one whole genome duplication in the history of vertebrates (Lynch & Conery, 2000). One of the most intensively studied cases of gene family expansion is that of Hox genes (Holland et al., 1994; Martinez & Amemiya, 2002; Ruddle et al., 1994b,a). Hox genes code for homeobox containing transcription factors which tend to occur in tightly linked clusters in the genome. They are involved in the development of vertebrate body plan characters (Shashikant et al., 1991). For that reason a causal connection between Hox gene family evolution and the evolution of the vertebrate body plan seems possible. In this paper we provide the first report on the Hox gene inventory of a species from the most primitive vertebrate lineage, the pacific hagfish *Eptatretus stoutii*.

The earliest vertebrates found in the fossil record are jawless (Janvier, 1996). Only two jawless groups are still alive, the lampreys and hagfish (Nelson, 1994). These lineages are considered recent offshoots of the large radiation of jawless vertebrates at the base of the vertebrate phylogeny (Forey & Janvier, 1993). Their phylogenetic affiliation, however, is unclear, both in terms of their relationship to the jawless groups of fossil vertebrates as well as their relationship to each other. On morphological grounds, lamprey and hagfish are often considered paraphyletic to jawed vertebrates (Fig. 1a) (Maisey, 1986). In contrast, most molecular phylogenies suggest that lampreys and hagfish form a monophyletic clade which is the sister taxon to the jawed vertebrates
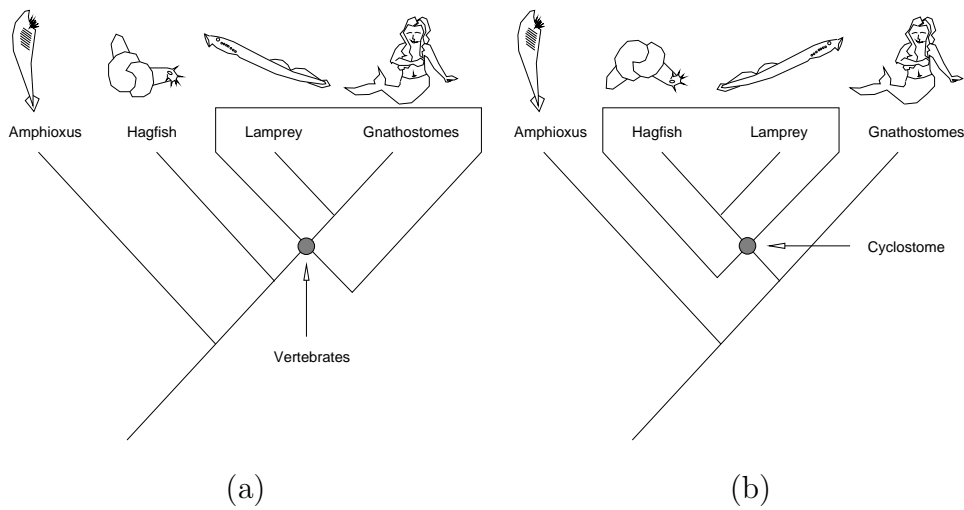
Fig. 1. Alternative phylogenies of early vertebrates. (a) The "vertebrate theory" assumes that lampreys and gnathostomes are a clade, the vertebrates, whose sister group is hagfishes. (b) The "cyclostome theory" assumes that lampreys and hagfishes are a clade, the cyclostomes, whose sister group is the gnathostomes.

(Fig. 1b) (Furlong & Holland, 2002; Kuraku et al., 1999; Mallatt & Sullivan, 1998; Stock & Whitt, 1992), with few dissenting voices (Gürsoy et al., 2000; Rasmussen et al., 1998). Hence, in this paper we interpret our data in the light of either possible phylogenetic hypothesis. The main conclusions in this paper will not depend on the difference between these two scenarios.

All vertebrate taxa examined today have been shown to contain multiple clusters of Hox genes. The best studied situation is that of the four mammalian Hox clusters, which are called *HoxA*, *HoxB*, *HoxC*, and *HoxD*, respectively (Kappen et al., 1989; Schughart et al., 1987). In contrast the sister taxon of vertebrates, the cephalochordates with amphioxus as a typical member, has a single Hox cluster with 14 different genes (Ferrier et al., 2000; Garcia-Fernández & Holland, 1994). Each vertebrate cluster has genes that can be assigned to a subset of 13 of the 14 genes found in amphioxus. Genes homologous to the same gene in the amphioxus Hox cluster are called a paralog group. There are thus 13 paralog groups among the mammalian Hox genes.

Four Hox clusters seem to be characteristic of all jawed vertebrates, with the notable exception of teleost fishes (Amores et al., 1998, 2003). These four Hox clusters are thought to have evolved in two rounds of whole genome duplication (Holland et al., 1994), one in the stem of all vertebrates and one in the stem lineage of jawed vertebrates, the gnathostomes. This scenario is also supported by the evolutionary history of non-Hox genes like *Engrailed* and *Otx* (Escriva et al., 2002; Germot et al., 2001; Holland & Williams, 1990). The situation at the base of the vertebrate phylogeny, however, is complicated by the fact that the two most basal lineages, lampreys and hagfish, are not well understood and also seem to have undergone independent gene and genome duplication

3

events (Escriva et al., 2002). For instance, it has been shown recently that the paralog Hox genes of lamprey have duplicated independently from those genes in the four mammalian clusters (Fried et al., 2003). Hence the lamprey data does not provide direct support to the 2R model which assumes a Hox cluster duplication prior to the most common recent ancestor of lampreys and jawed vertebrates. In this paper we show that the Hox gene inventory of hagfish, in contrast, does support the hypothesis of at least one round of Hox cluster duplication in the early history of the vertebrate stem. The picture is, however, also complicated by independent duplications and gene loss in the hagfish lineage.

2 Methods

DNA source

DNA was donated by Axel Meyer from a specimen of *Eptatretus stoutii* (pacific hagfish) collected by Rick Mayden (Alabama).

Genomic DNA surveys

Target sequences were amplified by PCR using the degenerate primer sets 5E/3F (Murtha et al., 1991) and 5E5/3F (Misof & Wagner, 1996; Misof et al., 1996):

5E	5'-AAAGGATTCCTGCAGARYTIGARAARGARTT-3'
5E5	5'-AAAGGATCCTGCAGAARMGITGYCCITAYASIAA-3'
3F	5'-ACAAGCTTGAATTCATICKICKRTTYTGRAACCA-3'

Primer set 5E/3F amplifies an 82-bp fragment from position 60 to 142 of the homeobox, yielding a 27-residue derived aminoacid sequence from position 20 to 47 of the homeodomain. The target sequences for primer set 5E/3F are homeoboxes in cognate groups 1-10. This primer set has been well tested on a variety of different organisms (Bartels et al., 1993; Cartwright et al., 1993; Dick & Buss, 1994; Murtha et al., 1991; Pendleton et al., 1993; Misof & Wagner, 1996; Ledje et al., 2002).

Primer set 5E5/3F amplifies a 114-bp fragment from position 27 to 141 of the homeobox, yielding a 38-residue derived aminoacid sequence from position 9 to 47 of the homeodomain. It proved to successfully amplify HOM/Hox

fragments of the posterior group 9 through 13 (Misof & Wagner, 1996; Ledje et al., 2002). It seems to consistently miss group 12 genes, however.

The PCR conditions for the experiments 5E/3F and 5E5/3F were identical to those reported in Misof & Wagner (1996). PCR conditions for the series 5E5-2/3F (with primers 5E5 and 3F) were 5 cycles (94°C, 1min; 37°C, 1min; 72°C, 1min) followed by 30 cycles (94°C, 1min; 50°C, 1min, 72°C, 1min) with varied $MgCl_2$ concentrations of 2.0mM, 2.5mM, and 3.0mM. Reactions were run with and without DMSO. PCR products from each reaction were individually cloned directly into pCR II vector (Invitrogen).

A saturation test for PCR surveys as described by Misof & Wagner (1996) was performed. The experimental series 5E5 and 5E5-2 appear to be saturated, the survey using the 5E/3F primer set deviates significantly from the exponential saturation model indicating that the primers are not unbiased.

The hagfish *Hox* sequences are available in Genbank with access numbers **AY445836**-**AY445871**.

Sequence Analysis

The sequences of the 278 PCR clones and their reverse complements were aligned using `ClustalW` and `dialign2` to identify identical sequences. In some cases sequences only differed from others by single point mutations, deletions, or duplications of single nucleotides, usually `G`. In such cases the consensus sequence of such a group of almost identical sequences is reported. The DNA sequences were then translated in all 6 possible reading frames and compared to known homeodomain proteins to obtain the putative Hox protein sequence fragments. Only those 41 distinct sequences that were unambiguously homologous to known Hox proteins were considered further.

`Blastn` searches in genbank were used to reconfirm that the sampled sequences were not contaminations by any known sequence. Contamination by human sequences can also be ruled out since none of the PCR products matched a known human Hox homeobox sequence.

In addition to 33-35 unique Hox genes, a putative *Gsx* homeodomeain protein and 12 unidentifyable sequences, probably PCR artifacts, were recovered. The assignment of the 33 distinct sequences to paralog groups was done by direct comparison of the inferred protein sequences with the corresponding sequences from human, latimeria, shark, and lamprey. A neighborjoining tree summarizing these data is shown in Fig. 2.
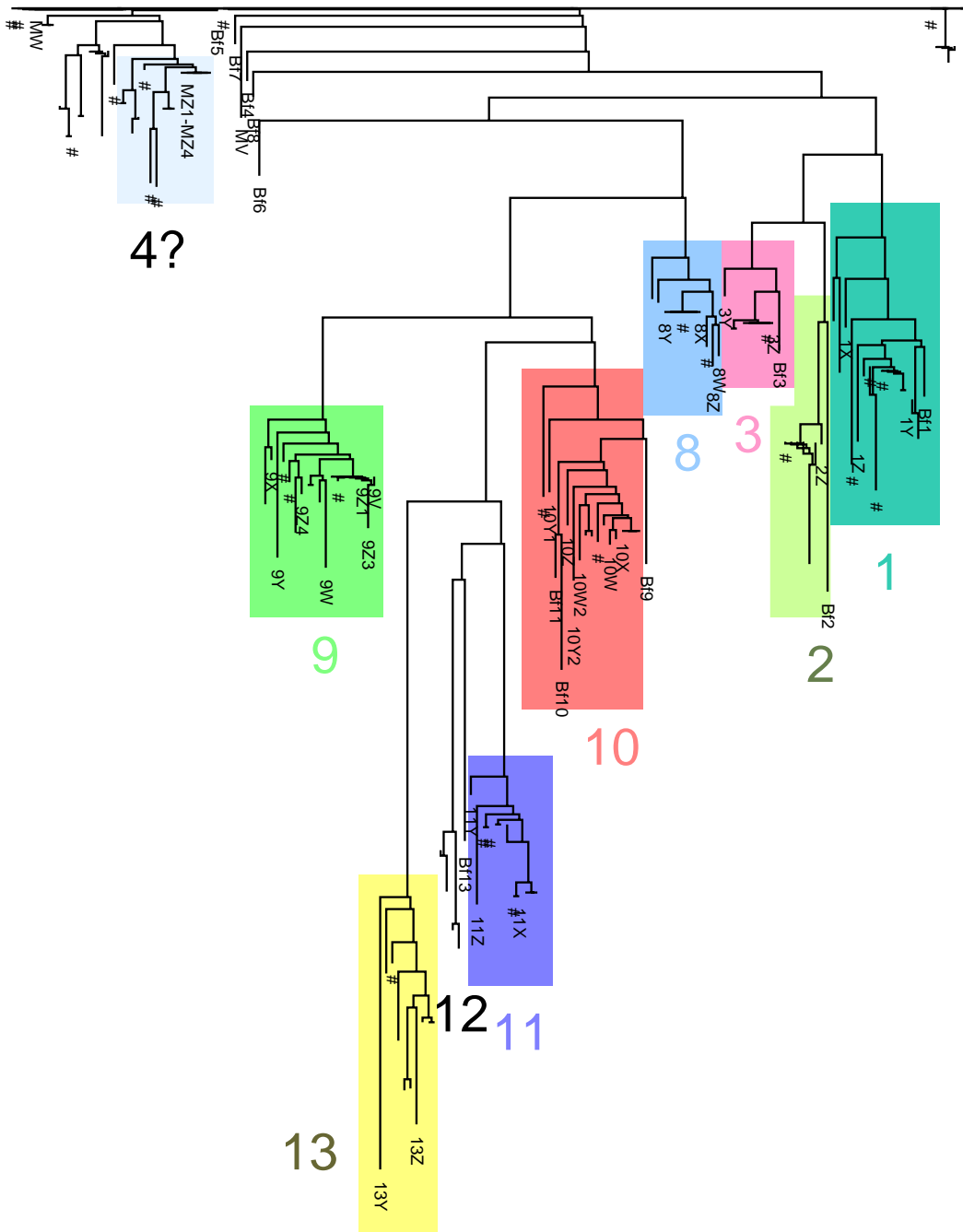
Fig. 2. Neighborjoining tree of the partial protein sequences of all 33 hagfish Hox sequences from this study (labeled) and sequences from amphioxus (Bf-number), lamprey (#), and gnathostomes (shark, latimeria, human; no label). The identification of HoxMZ1 through HoxMZ4 as paralog group 4 genes is only tentative.

As the phylogenetic signals in the short sequences are weak both at the protein and the nucleic acid level detailed analysis of the relationships with a paralog group and the resolution of the middle group Hox4-Hox7 genes proved impossible. In Fried et al. (2003) it is shown that the lamprey Hox genes typically cluster outside the gnathostome Hox genes of the same paralog group. The amphioxus sequences typically do not cluster well with the gnathostome sequences (Ferrier et al., 2000). We have therefore employed the quartet mapping technique (Nieselt-Struwe & von Haeseler, 2001) in order to detect similarities of the individual hagfish Hox sequences with either the gnathostomes, lampreys, or amphioxus sequences. This method requires that sequences are partitioned into four groups, P, Q, R, S, in our case a hagfish sequence, an amphioxus sequence, and a sample of lamprey and gnathostome sequences, respectively. For each quadruple of sequences $p \in P$, $q \in Q$, $r \in R$, $s \in S$ the number n_i of columns in the alignment that support one of the three possible trees $\mathfrak{T}_1 = (PQ)(RS)$, $\mathfrak{T}_2 = (PR)(QS)$, and $\mathfrak{T}_3 = (PS)(QR)$ is computed. The fractions $x_i = n_i/(n_1 + n_2 + n_3)$ can then be represented in a unit simplex. The distribution of the points for all quadruples summarizes the phylogenetic relationships between the four groups; see Fig. 3.

In addition to quartet mapping we have attempted to reconstruct trees for each paralog group using neighbor joining (Saitou & Nei, 1987), maximum parsimony, and maximum likelihood as implemented in J. Felsenstein's `phylip` package (Felsenstein, 1989), as well as canonical split decomposition (Bandelt & Dress, 1992) as implemented in `splitstree 3.1` (Huson, 1998). In general, the signal is too weak to reconstruct reliable trees from the protein sequences, from nucleic acid sequences, and from nucleic acid sequences restricted to the 1st and 2nd codon positions. Even retaining all internal nodes with a bootstrap support of $\geq 30\%$ yields very poorly resolved trees, see electronic supplemental material.

3   Results and Discussion

We amplified homeobox fragments from genomic DNA of the pacific hagfish, *Eptatretus stoutii*, using two primer sets. 5E/3F is targeting Hox genes from paralog groups 1 to 10 (Murtha et al., 1991), while 5E5/3F targets paralog groups 9 to 13 (Misof & Wagner, 1996). Since the yield of AbdB related Hox sequences with the primer 5E5-3F under standard conditions was low we repeated the PCR experiment with varying Mg concentrations (Misof & Wagner, 1996). The results of this sample are called 5E5-2. A total of 278 clones were sequenced, 167 from an amplification with 5E/3F and 33 and 78, respectively,
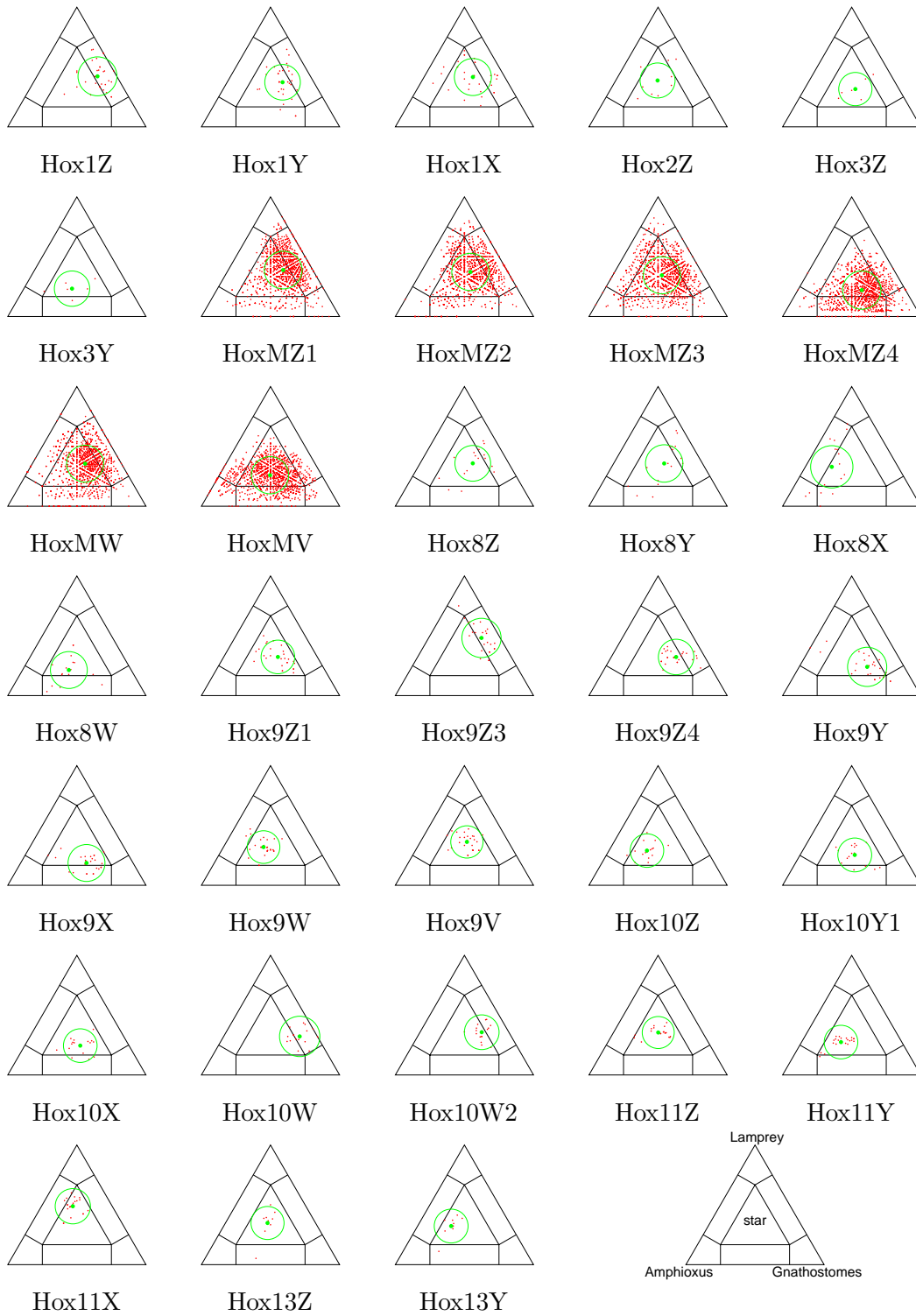
Fig. 3. Quartet mapping (Nieselt-Struwe & von Haeseler, 2001) of the 33 Hagfish Hox sequences in comparison with amphioxus, lamprey, and gnathostome sequences from the same paralog group. The circles indicate the mean fraction (center) and the standard deviation (radius).

8

Table 1

Saturation tests. Here n is the number of PCR clones containing a Hox gene, N_{found} is the number of distinct sequences (not counting allelic variants but treating HoxMY, HoxMW, and HoxMU2 as distinct genes. N_{exp} is the expected number of distinct Hox genes from the first order saturation kinetics model, and p_{miss} is the probability of missing a gene in a sample of size n, see Misof & Wagner (1996) and the electronic supplement for details.

PCR series	n	N_{found}	N_{exp}	p_{miss}
5E5/3F	29	8	8.02	0.021
5E5-2/3F	64	14	14.52	0.010
5E/3F	144	24	22.88	0.002

from amplifications with 5E5/3F under two sets of PCR conditions. Of these, 237 (144, 29, and 64) were Hox sequences, and 9 corresponded to a single *Gbx* parahox gene.

The two 5E5/3F clone sets contained eight and fourteen unique sequences respectively. Saturation analysis (Misof & Wagner, 1996) indicates that the set of unique sequences is likely to be complete. The predicted number of unique sequences from the saturation model matches closely the number of unique sequences found in the experiments 5E5/3F and 5E52/3F (see Table 1). The saturation model does not match as well the data from the 5E/3F amplification as it does from the 5E5/3F and 5E52/3F amplifications. In the case of 5E/3F, the number of unique clones found, 24, exceeds that estimated by the model. A similar situation was noted in Misof & Wagner (1996); Misof et al. (1996) in studies of the zebrafish and killifish Hox genes. This is most likely due to heterogeneity in the probability of different paralog groups to amplify with these primers and thus violate the assumptions of the saturation model. Overall, however, the sample of cloned sequences is likely to represent all the sequences amplified. This result, of course, does not exclude the possibility that there are additional Hox genes in the hagfish genome that were not amplified in our reactions.

A test for allelic variants was applied as described in Misof & Wagner (1996). From the originally 41 distinct Hox sequences 33 are identified as unambiguously distinct Hox fragments. Three middle group sequences, HoxMY and HoxMW, and HoxMU2 have pairwise Hamming distances of 12, compared to the threshold value of 11 for the allele test. We conservatively treat them as a single gene in the subsequent analysis.

Our PCR survey provides evidence for at least 33 different Hox genes in the hagfish genome. The homebox sequence fragments can unambiguously be clas-
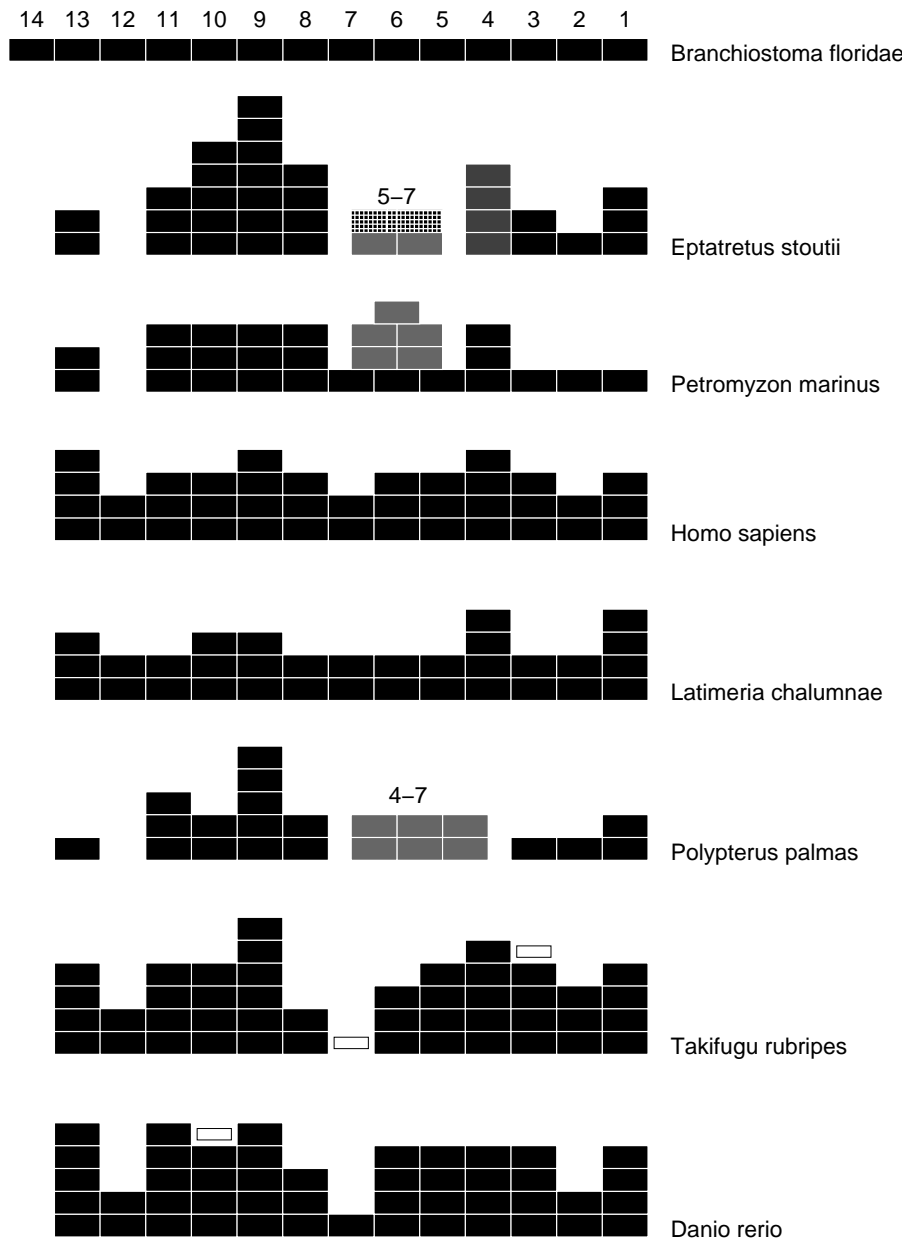
Fig. 4. Comparison of Hox gene inventory. Pseudogenes are shown as open boxes, genes for which the paralog group cannot be determined unambiguously are shown in grey. Two middle groups genes in the hagfish are marked differently to indicate that there are 2 or 4 genes in groups 5-7 depending on whether HoxMW, MoxMY, and HoxMU2 are distinct genes or allelic variants. Except for *Homo sapiens*, where a completed genome is available, all numbers are lower bounds.

Data sources: *Branchiostoma floridae* (Ferrier et al., 2000), *Eptatretus stoutii* this study, *Petromyzon marinus* (Force et al., 2002; Irvine et al., 2002), *Latimeria chalumnae* (Koh et al., 2003), *Homo sapiens* Genbank; *Takifugu rubripes* (Amores et al., 2003), *Polypterus palmus* (Ledje et al., 2002).

Table 2
Summary of support for the different phylogenetic positions of Hagfish Hox genes
given in Fig. 3. H ... hagfish, G ... gnathostomes (represented by human, latimeria,
and shark), L ... lamprey, A ... amphioxus.

Tree	Mean	sdv
(H,A)(G,L)	0.306	0.016
(H,G)(A,L)	0.367	0.013
(H,L)(G,A)	0.327	0.012

sified into paralog groups, with the exception of medial group genes and genes
for group four. The number of homeobox fragments identified per paralog
group, to the exclusion of medial group genes, is very variable and ranges
from zero in group 12 to seven in group 9. This data is most consistent with
the hypothesis of multiple clusters, but in itself can not provide evidence
for physical linkage between the genes. Hence we refrain from predicting the
number of Hox clusters. Comparison with other chordate Hox gene inventories
reveals a considerable over-representation of Hox-9 group fragments (Fig. 4).
The highest number of group 9 paralogs have been described in actinoptery-
gian fishes: zebrafish, 5 (Amores et al., 1998), killifish, 5 (Misof & Wagner,
1996), and pufferfish, 6 (Amores et al., 2003), as well as bichir, 5 (Ledje et al.,
2002). Considering the phylogenetic position of hagfish and actinopterygians,
the hagfish Hox gene inventory suggest that Hox gene duplications happened
in the hagfish lineage independently of the rest of vertebrates. This conclusion
is also supported by the analysis of gene phylogenies of zebrafish and fugu Hox
genes (Amores et al., 1998, 2003)), which show that the additional ray finned
fish Hox genes are due to a duplication within the actinopterygian lineage.

Not surprisingly the phylogenetic signal in the small homeobox fragments is
weak and hence a detailed gene phylogeny can not be reconstructed. Never-
theless, in some cases a clear signal can be obtained which sheds light on the
genomic history of early vertebrates. First, two gene pairs, Hox8X and Hox8Y
as well Hox9Z1 and Hox9V are most likely first order paralogs, supporting
the conclusion that independent Hox cluster duplications must have taken
place in the hagfish lineage (see previous paragraph). This situation is simi-
lar to the findings about lamprey Hox genes, which, within a paralog group,
are predominantly (Pendleton et al., 1993; Sharman & Holland, 1998; Force
et al., 2002; Irvine et al., 2002) or perhaps even exclusively derived within the
lamprey lineage (Fried et al., 2003). Hence it is possible that no Hox cluster
duplication may have happened in the stem lineage of vertebrates, i.e. prior
to the most recent common ancestor of Recent jawless and jawed vertebrates.

To address the question whether the first Hox cluster duplication happened

before or after the most recent common ancestor of gnathosomes and hagfish we performed a Quartet Mapping analysis of the support for the three possible gene trees among amphioxus, lamprey, hagfish and gnathostome Hox genes (Fig. 3, Tab. 2). There is no support for the unrooted tree

$$((\text{Hagfish},\text{Amphioxus})(\text{Lamprey},\text{Gnathosome})) = ((\text{H},\text{A})(\text{L},\text{G})),$$

i.e. the mean support level is very closed to 1/3, as expected for a perfect star phylogeny. The mean support for ((H,L)(A,G)) tree is about 2.25 standard deviations higher than 1/3 and may reflect some gene duplications which are shared between hagfish and lamprey, but not with gnathosomes. By far the strongest support overall, more than five standard deviations larger than 1/3, is for the unrooted tree which associates some of the hagfish genes with gnathostome Hox genes: ((H,G)(A,L)). The weak phylogenetic signal from the tree reconstruction methods is consistent with this picture. Note that the purpose of these estimates is not to reconstruct the phylogeny of the basal recent chordate taxa, but to find the closest relatives of paralog genes, given our current understanding of vertebrate phylogeny.

There are only two rooted phylogenies for the major recent chordate taxa currently under discussion: (O(A(H(L,G)))), and (O(A((H,L)G))). Molecular phylogenies tend to support the monophyletic status of a hagfish/lamprey clade (Forey & Janvier, 1993; Delabre et al., 2002; Furlong & Holland, 2002; Takezaki et al., 2003), however. Either phylogenetic hypothesis together with the Quartet Mapping data and the previously published results for lamprey (Fried et al., 2003) have two implications. First they suggest that there was more than one Hox cluster in the most recent common ancestor of gnathostomes and cyclostomes. This conclusion is inferred from the significant association between hagfish Hox genes with gnathosome Hox genes. This interpretation is consistent with the 2R hypothesis of Holland et al. (1994). Second, the results imply that the Hox gene situation in lamprey must be highly derived with many Hox gene lineages extinct and replaced with more recent duplicates. These conclusions are consistent with other well supported facts:

(1) There is a single Hox cluster in the amphioxus *Branchiostoma floridae* (Garcia-Fernández & Holland, 1994; Ferrier et al., 2000).
(2) All gnathostomes investigated so far with the exception of higher teleosts have four Hox clusters that are homologous and most likely derived by duplication from the single amphioxus Hox cluster (Holland & Garcia-Fernandez, 1996; Koh et al., 2003; Prohaska et al., 2003b).
(3) The higher teleosts underwent an additional duplication of the four gnathostome clusters (Amores et al., 1998; Stellwag, 1999; Chiu et al., 2002; Málaga-Trillo & Meyer, 2001; Prohaska et al., 2003a; Amores et al., 2003).
(4) Lampreys have three to four clusters that have no clear homology to either individual or pairs of gnathostome clusters (Pendleton et al., 1993;
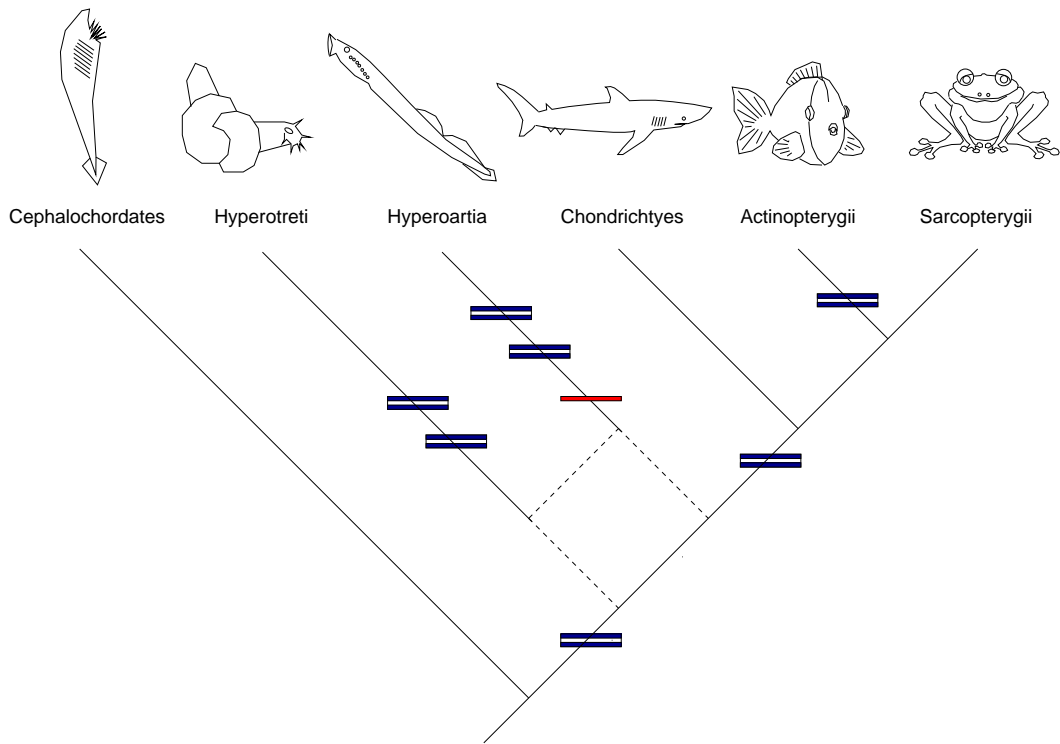
Fig. 5. A plausible history of Hox cluster duplications that is consistent with the 2R model. For details see text. Duplication events are indicated by =, the single bar indicates the loss of one gene in (almost) all paralog groups in the lamprey lineage.

Sharman & Holland, 1998; Force et al., 2002; Irvine et al., 2002; Fried et al., 2003).

A plausible scenario for the evolution of the Hox gene data, Fig. 5 runs as follows:

(1) The first duplication of the cephalochordate Hox cluster occurred in a common ancestor of gnathostomes, lampreys and hagfishes. The redundancy introduced in the duplication was resolved differently in these three groups.
(2) While hagfishes and gnathostomes retained a large number of first order paralogs, the redundant genes were essentially lost in the lamprey lineage.
(3) A further round of duplication lead to the four gnathostome clusters.
(4) Two independent rounds of duplications occurred in both the lamprey and hagfish lineages.

This model explains a number of findings: (i) The fact that the lamprey Hox sequences seem to arise from a single cluster (Fried et al., 2003). (ii) The fact that so far it has not been possible to determine the linkage groups of all lamprey Hox clusters. In our present model they should be located in

up to 8 different groups with none of them spanning the entire range from PG1 to PG13. One might speculate that, after the first duplication, genes in approximately PG 4 to PG 9 or 10 were retained on one copy while both "ends" were located on the other copy. This would explain the known pattern of closely linked genes in lampreys and the failure to assemble complete clusters by "genome walking" in (Force et al., 2002; Irvine et al., 2002). (iii) The fact that hagfish and gnathostome sequences seem to be related more closely than lampreys and gnathostomes or hagfish and amphioxus (because hagfish and gnathostome Hox sequences have retained a weak signal from the duplication event of their common ancestor).

Acknowledgements

References

Amores, A., Force, A., Yan, Y. L., Joly, L., Amemiya, C., Fritz, A., Ho, R. K., Langeland, J., Prince, V., Wang, Y. L., Westerfield, M., Ekker, M., Postlethwait, J. H., 1998. Zebrafish hox clusters and vertebrate genome evolution. Science 282, 1711–1714.

Amores, A., Suzuki, T., Yan, Y.-L., Pomeroy, J., Singer, A., Amemiya, C., Postlethwait, J., 2003. Developmental roles of pufferfish *Hox* clusters and genome evolution in ray-fin fish. Genome Res. In press.

Bandelt, H. J., Dress, A. W. M., 1992. A canonical decomposition theory for metrics on a finite set. Adv. Math. 92, 47.

Bartels, J. L., Murtha, M. T., Ruddle, F. H., 1993. Multiple Hox/HOM-class homeoboxes in plathelminthes. Mol. Phylog. Evol. 2, 143–151.

Cartwright, P., Dick, M. H., Buss, L. W., 1993. HOM/Hox type homeoboxes in the chelicerate *Limulus polyphemus*. Mol. Phylog. Evol. 2, 185–192.

Chiu, C.-h., Amemiya, C., Dewar, K., Kim, C.-B., Ruddle, F. H., Wagner, G. P., 2002. Molecular evolution of the HoxA cluster in the three major gnathostome lineages. Proc. Natl. Acad. Sci. USA 99, 5492–5497.

Delabre, C., Gallut, C., Barriel, V., Janvier, P., Gachelin, G., 2002. Complete mitochondrialDNA of the hagfish, *Eptatretus burgerie*: The comparative anaylsis of mitochondrial DNA sequences strongly supports cyclostome monophyly. Mol. Phylog. Evol. 22, 184–192.

Dick, M. H., Buss, L. W., 1994. A PCR-based survey of homeobox genes in *Ctenodrilus serratus* (annelida: Polychaeta). Mol. Phylog. Evol. 3, 146–158.

Escriva, H., Manzon, L., Youson, J., Laudet, V., 2002. Analysis of lamprey and hagfish genes reveals a complex history of gene duplications during early vertebrate evolution. Mol. Biol. Evol. 19, 1440–1450.

Felsenstein, J., 1989. Phylip – phylogeny inference package (version 3.2). Cladistics 5, 164–166.

Ferrier, D. E. K., Minguillón, C., Holland, P. W. H., Garcia-Fernàndez, J., 2000. The amphioxus Hox cluster: deuterostome posterior flexibility and *Hox14*. Evol. Dev. 2, 284–293.

Force, A., Amores, A., Postlethwait, J. H., 2002. Hox cluster organization in the jawless vertebrate *Petromyzon marinus*. J. Exp. Zool. (Mol. Dev. Evol.) 294, 30–46.

Forey, P., Janvier, P., 1993. Agnathans and the origin of jawed vertebrates. Nature 361, 129–134.

Fried, C., Prohaska, S. J., Stadler, P. F., 2003. Independent Hox-cluster duplications in lampreys. J. Exp. Zool., Mol. Dev. Evol 299B, 18–25.

Furlong, R. F., Holland, P. W., 2002. Bayesian phylogenetic analysis supports monophyly of ambulacria and cyclostomes. Zool. Sci, 19, 593–599.

Garcia-Fernández, J., Holland, P. W., 1994. Archetypal organization of the amphioxus hox gene cluster. Nature 370, 563–566.

Germot, A., Lecointre, G., Plouhinec, J.-L., Le Mentec, C., Girardot, F., Mazan, S., 2001. Structural evolution of *otx* genes in craniates. Mol. Biol. Evol. 18, 1668–1678.

Gürsoy, H.-C., Koper, D., Benecke, B.-J., 2000. The vertebrate 7S K RNA separates hagfish (*Myxine glutinosa*) and lamprey (*Lampetra fluviatilis*). J. Mol. Evol. 50, 456–464.

Holland, P. W., Garcia-Fernandez, J., 1996. Hox genes and chordate evolution. Dev. Biol. 173, 382–395.

Holland, P. W., Williams, N. A., 1990. Conservation of engrailed-like homeobox sequences during vertebrate evolution. FEBS Lett. 277, 250–252.

Holland, P. W. H., Garcia-Fernández, J., Williams, N. A., Sidow, A., 1994. Gene duplication and the origins of vertebrate development. Development (Suppl.), 125–133.

Huson, D. H., 1998. Splitstree: analyzing and visualizing evolutionary data. Bioinformatics 14, 68–73.

Irvine, S. Q., Carr, J. L., Bailey, W. J., Kawasaki, K., Shimizu, N., Amemiya, C. T., Ruddle, F. H., 2002. Genomic analysis of Hox clusters in the sea lamprey, *Petromyzon marinus*. J. Exp. Zool. (Mol. Dev. Evol.) 294, 47–62.

Janvier, P., 1996. Early Vertebrates. Clarendon Press, Oxford.

Kappen, C., Schughart, K., Ruddle, F. H., 1989. Two steps in the evolution of antennapedia-class vertebrate homeobox genes. Proc. Natl. Acad. Sci. USA 86, 5459–5463.

Koh, E. G. L., Lam, K., Christoffels, A., Erdmann, M. V., Brenner, S., Venkatesh, B., 2003. *Hox* gene clusters in the indonesian coelacanth, *Latimeria menadoensis*. Proc. Natl. Acad. Sci. USA 100, 1084–1088.

Kuraku, S., Hoshiyama, D. Katoh, K., Suga, H., Miyata, T., 1999. Monophyly

of lampreys and hagfishes supported by nuclear DNA-coded sequences. J. Mol. Evol. 49, 729–735.

Ledje, C., Kim, C.-B., H., R. F., 2002. Characterization of Hox genes in the bichir, *Polypterus palmas*. J. Exp. Zool., Mol. Dev. Evol. 294, 107–111.

Lynch, M., Conery, J. S., 2000. The evolutionary fate and consequences of duplicate genes. Science 290, 1151–1155.

Maisey, J. G., 1986. Heads and tails: a chordate phylogeny. Cladistics 2, 201–256.

Málaga-Trillo, E., Meyer, A., 2001. Genome duplications and accelerated evolution of *Hox* genes and cluster architecture in teleost fishes. Amer. Zool. 41, 676–686.

Mallatt, J., Sullivan, J., 1998. 28S and 18S rDNA sequences support the monophyly of lampreys and hagfishes. Mol. Biol. Evol. 15, 1706–1718.

Martinez, P., Amemiya, C. T., 2002. Genomics of the HOX gene cluster. Comp. Biochem. Physiol. Part B 133, 571–580.

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, B. Y., Blanco, M. J., Wagner, G. P., 1996. A PCR-survey of *Hox* genes of the zebrafish: new sequences and evolutionary implications. J. Exp. Zool. 274, 193–206.

Misof, B. Y., Wagner, G. P., 1996. Evidence for four Hox clusters in the killifish *Fundulus Heteroclitus* (teleostei). Mol. Phylog. Evol. 5, 309–322.

Murtha, M. T., Lechman, J. F., Ruddle, F. H., 1991. Detection of homeobox genes in development and evolution. Proc. Natl. Acad. Sci. USA 88, 10711–10715.

Nelson, J. S., 1994. Fishes of the World. John Wiley & Sons, Inc., New York.

Nieselt-Struwe, K., von Haeseler, A., 2001. Quartet-mapping, a generalization of the likelihood mapping procedure. Mol. Biol. Evol. 18, 1204–1219.

Ohno, S., 1970. Evolution by Gene Duplication. Springer Verlag, New York.

Pendleton, J., Nagai, B. K., Murtha, M. T., Ruddle, F. H., 1993. Expansion of the Hox gene family and the evolution of chordates. Proc. Natl. Acad. Sci. USA 90, 6300–6304.

Prohaska, S., Fried, C., Flamm, C., Wagner, G., Stadler, P. F., 2003a. Surveying phylogenetic footprints in large gene clusters: Applications to Hox cluster duplications. Mol. Phyl. Evol. In press; doi:10.1016/j.ympev.2003.08.009.

Prohaska, S. J., Fried, C., Amemiya, C. T., Ruddle, F. H., Wagner, G. P., Stadler, P. F., 2003b. The shark HoxN cluster is homologous to the human HoxD cluster. J. Mol. Evol. In press.

Rasmussen, A.-S., Janke, A., Arnason, U., 1998. The mitochondrial DNA molecule of the hagfish *Myxine glutinosa* and vertebrate phylogeny. J. Mol. Evol. 46, 382–388.

Ruddle, F. H., Bartels, J. L., Bentley, K. L., Kappen, C., Murta, M. T., W., P. J., 1994a. Evolution of Hox genes. Annu. Rev. Genet. 28, 423–442.

Ruddle, F. H., Bentley, K. L., Murtha, M. T., Risch, N., 1994b. Gene loss and gain in the evolution of the vertebrates. Development (Supplement), 155–161.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol. Evol. 4, 406–425.

Schughart, K., Kappen, C., Ruddle, F. H., 1987. Mammalian homeobox-containing genes: genomic organization, structure, expression and evolution. Brit. J. Cancer 58, 9–13.

Sharman, A. C., Holland, P. W., 1998. Estimation of Hox gene cluster number in lampreys. Int. J. Dev. Biol. 42, 617–620.

Shashikant, C. S., Utset, M. F., Violette, S. M. Wise, T. L., Einat, P., Einat, M., Pendleton, J. W., Schughart, K., Ruddle, F. H., 1991. Homeobox genes in mouse development. Crit. Rev. Eukaryot. Gene Expr. 1, 207–245.

Stellwag, E. J., 1999. Hox gene duplications in fish. Cell Devel. Biol. 10, 531–540.

Stock, D. W., Whitt, G. S., 1992. Evidence from 18S ribosomal RNA sequences that lampreys and hagfish form a natural group. Science 257, 787–789.

Takezaki, N., Figueroa, F., Zelska-Rutcynska, Z., Klein, J., 2003. Molecular phylogeny of early vertebrates: Monophyly of the agnathans as revealed by sequences of 35 genes. Mol. Biol. Evol 20, 287–292.