

Research article

Open Access

Phylogenetic and chromosomal analyses of multiple gene families syntenic with vertebrate Hox clusters

Görel Sundström, Tomas A Larsson and Dan Larhammar*

Address: Department of Neuroscience, Uppsala University, Box 593, 75124 Uppsala, Sweden

Email: Görel Sundström - Gorel.Sundstrom@neuro.uu.se; Tomas A Larsson - Tomas.Larsson@neuro.uu.se; Dan Larhammar* - Dan.Larhammar@neuro.uu.se

* Corresponding author

Published: 19 September 2008

Received: 11 March 2008

BMC Evolutionary Biology 2008, **8**:254 doi:10.1186/1471-2148-8-254

Accepted: 19 September 2008

This article is available from: <http://www.biomedcentral.com/1471-2148/8/254>

© 2008 Sundström et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Ever since the theory about two rounds of genome duplication (2R) in the vertebrate lineage was proposed, the Hox gene clusters have served as the prime example of quadruplicate paralogy in mammalian genomes. In teleost fishes, the observation of additional Hox clusters absent in other vertebrate lineages suggested a third tetraploidization (3R). Because the Hox clusters occupy a quite limited part of each chromosome, and are special in having position-dependent regulation within the multi-gene cluster, studies of syntenic gene families are needed to determine the extent of the duplicated chromosome segments. We have analyzed in detail 14 gene families that are syntenic with the Hox clusters to see if their phylogenies are compatible with the Hox duplications and the 2R/3R scenario. Our starting point was the gene family for the NPY family of peptides located near the Hox clusters in the pufferfish *Takifugu rubripes*, the zebrafish *Danio rerio*, and human.

Results: Seven of the gene families have members on at least three of the human Hox chromosomes and two families are present on all four. Using both neighbor-joining and quartet-puzzling maximum likelihood methods we found that 13 families have a phylogeny that supports duplications coinciding with the Hox cluster duplications. One additional family also has a topology consistent with 2R but due to lack of urochordate or cephalocordate sequences the time window when these duplications could have occurred is wider. All but two gene families also show teleost-specific duplicates.

Conclusion: Based on this analysis we conclude that the Hox cluster duplications involved a large number of adjacent gene families, supporting expansion of these families in the 2R, as well as in the teleost 3R tetraploidization. The gene duplicates presumably provided raw material in early vertebrate evolution for neofunctionalization and subfunctionalization.

Background

The presence of several paralogous gene regions, forming so-called paralogs [1] with gene family members on two to four chromosomes in mammals, has been taken as evidence for two rounds of whole genome duplication

"2R", deduced to have taken place in the ancestor of vertebrates [1-11]. The Hox gene clusters (located on human chromosomes 2, 7, 12 and 17) have been used as the prime example of quadruplicate paralogy in vertebrate genomes, reviewed by Hoegg and Meyer [12], as com-

pared to the single Hox cluster in the cephalochordate lineage [13]. Until recently the number of Hox clusters in cartilaginous fishes has been unclear. The sequencing of the elephant shark genome (*Callorhynchus milii*) established that four Hox-clusters were already present in the ancestor of jawed vertebrates [14]. The Hox clusters are also known to be special in many ways, particularly in the way they are transcriptionally regulated [15]. Tunicate species like *Ciona intestinalis* and *Oikopleura dioika*, which have broken Hox clusters [16-18], are considered to be exceptions from the linear organization of vertebrate and cephalochordate Hox clusters, because the linear cluster of Hox genes was present in the last common ancestor of all bilaterians and possibly before the origin of Cnidarians [15,19].

In ray-finned fishes, an additional tetraploidization (3R) has been inferred based on the occurrence additional Hox clusters not present in other vertebrate lineages [12,20-22]. This has been confirmed using whole genome data from fully sequenced teleosts [23-27]. The study of Hox genes as well as other gene families in several species of actinopterygians has further indicated that the duplication event took place early in the evolution of actinopterygians [22,28]. It has also been proposed that the high number of duplicate genes created in 3R was in part involved in the radiation of teleost fishes [28-30]. The great number of species within the Euteleostei as compared to more basal actinopterygians supports the idea that genome duplication and speciation are linked [24,28] even though there seem to be a large time span between the duplication event and the actual radiation of euteleosts [31,32]. Several teleost lineages, for example Cyprinidae, Salmonidae, and Catostomidae [33-36] are known polyploids of relatively recent origin. The relation between polyploidy and speciation is not yet fully clear but it is worth noting that several species-rich fish clades are polyploids [37]. Both the existence of tetraploids and evidence of independent gene duplications indicate that fish genomes are in some sense more "plastic" than genomes in other vertebrates [29].

After duplication, the genes can undergo nonfunctionalization, subfunctionalization or neofunctionalization. It has been calculated that a duplicated gene in general has an average half-life of approximately 4 million years before it will be silenced [38]. But there are also examples of genes that have recently become pseudogenes even if the duplication took place a long time ago. Some of those examples are the Hoxb7 gene in *Takifugu rubripes* [39] and many of the olfactory genes in hominids [40]. Another example is the neuropeptide Y receptor Y6 that arose very early in vertebrate evolution and later has been mutated to become a pseudogene seemingly independently in many mammalian lineages [41]. The rate of retention of genes

after duplication has been investigated in several lineages. The obtained values vary considerably depending on species and time since tetraploidization, for example 8–16% in yeast [42-44], 15–24% for teleosts after 3R [25,45,46], 30% in *Arabidopsis* after its last tetraploidization [47] and about 50% after the salmonid tetraploidization [30]. It also seems like there is (at least in *Arabidopsis*) a connection between "survival" after the first and any subsequent genome duplications. Genes that have been preserved after the first tetraploidization are more likely to be retained after subsequent tetraploidization [48]. However, several factors obscure analyses of old duplication events, such as the frequent loss of genes after duplication, gene conversion and unequal evolutionary rates of paralogs. Because of this the use of the topology of phylogenetic trees as the only criterion for inferring block duplications is not sufficient and the symmetrical (A, B), (C, D) topology expected by two duplication events is rarely seen [9,49]. We believe that a combination of phylogenetic and map-based approaches using several species is much more useful to resolve the evolutionary history of gene families as has been pointed out earlier [50,51]. Although there is evidence for at least one tetraploidization in early vertebrate evolution the concept of two basal vertebrate tetraploidizations has been difficult to establish until recent studies that used both phylogenetic and positional information [11] or only the latter [10] from several sequenced vertebrate genomes. The recent publication of the amphioxus genome confirmed quadruplication of gene regions in gnathostomes [52].

Because each Hox cluster is located in a rather limited part of its chromosome, studies of more gene families are needed to fully understand the evolutionary history of the Hox-bearing chromosomes. In this work we have performed phylogenetic analyses of gene families located near the Hox clusters in several vertebrates in order to see if there are additional gene families with an evolutionary history similar to that of the Hox clusters. Our starting point in this work was the position of the genes coding for neuropeptide Y (NPY) family peptides in the pufferfish *Takifugu rubripes* and the zebrafish *Danio rerio* compared to the positions in the human genome.

Results

In total, 14 protein families predicted by the Ensembl database were studied in detail (see Fig. 1A–D and Additional files 1 and 2). Some of these families were further divided into subfamilies based on positions of invertebrate sequences in the initial phylogenetic analysis. Species included in the initial analysis were chosen in order to achieve relative dating of the duplications, thus avoiding analyses of duplication events that occurred before the emergence of chordates (see methods).

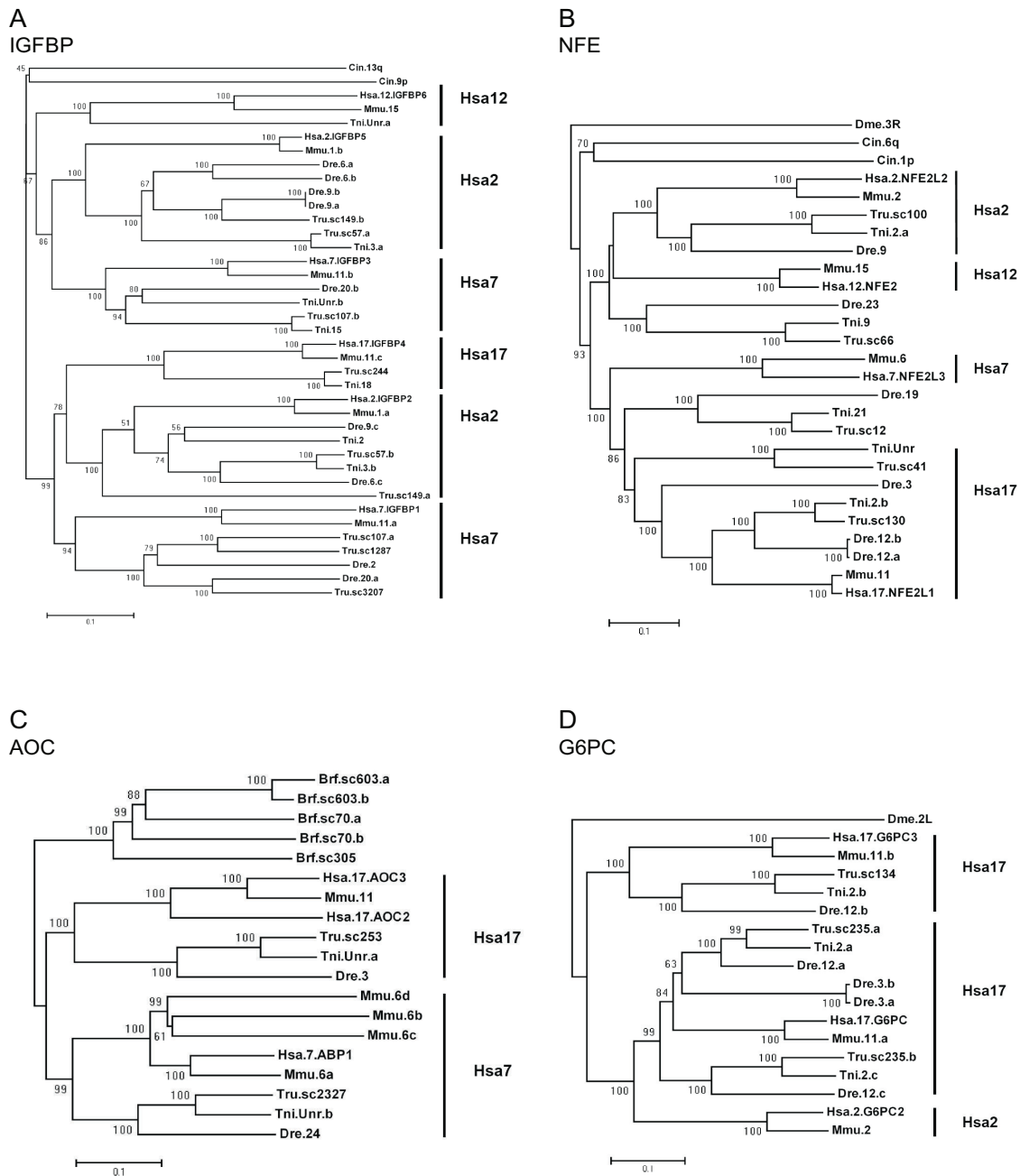


Figure 1
A-D – Phylogenetic trees for four of the investigated gene families. Phylogenetic neighbor-joining trees for four of the gene families analyzed in this study (A: IGFBP, B: NFE, C: AOC and D: G6PC). All trees constructed with the NJ method as implemented in Clustal W 1.81 with 1000 bootstrap replicates (values shown for each node). For IGFBP, sequences without the domains PF00219 (Insulin-like growth factor binding protein) and PF00086 (Thyroglobulin type-I repeat) were removed from the alignment. Note that the tree contains two subfamilies relevant for this paralogon. Hsa is human (*Homo sapiens*), Mmu: Mouse (*Mus musculus*), Dre: zebrafish (*Danio rerio*), Tru: fugu (*Takifugu rubripes*), Tni: tetraodon (*Tetraodon nigroviridis*), Cin: (*Ciona intestinalis*), Brf (*Branchiostoma floridae*) and Dme: (*Drosophila melanogaster*). For complete phylogenetic analysis of all 14 families see Additional files 1 and 2.

Three families in addition to the 14 analyzed families fulfilled the selection criteria (see methods) but had to be excluded for technical reasons, namely because repeated domains made the sequences difficult to align, thereby precluding reliable phylogenetic analyses. These families were ASB (ankyrin-repeat proteins with a SOCS box), HDAC (histone deacetylase) and MGAT (mannoside acetylglucosaminyltransferase).

For some of the fish sequences no clear human orthologs are found in the phylogenetic trees (one such example is the NR1D tree where the fish genes are clustered together

without any tetrapod sequence). However, when chromosome position information is taken into account it becomes apparent that there is conserved synteny between fish and human chromosomes indicated by striped boxes in Figs. 2, 3, 4, 5. For a majority of the phylogenetic trees the topologies in the NJ and ML analysis are in agreement (see Additional files 1 and 2).

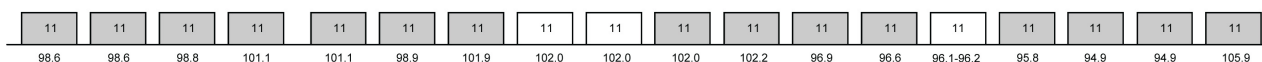
Gene families represented on four human Hox chromosomes

IGFBP, insulin like growth factor binding proteins, are part of a loosely defined superfamily [53]. The IGFBP fam-

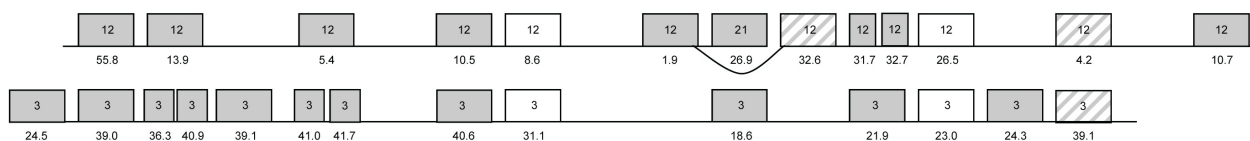
Hsa 17



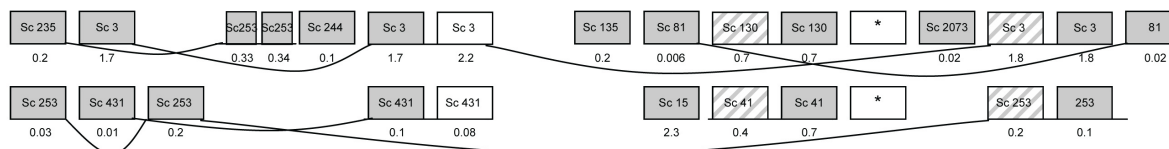
Mmu



Dre



Tru



Tni

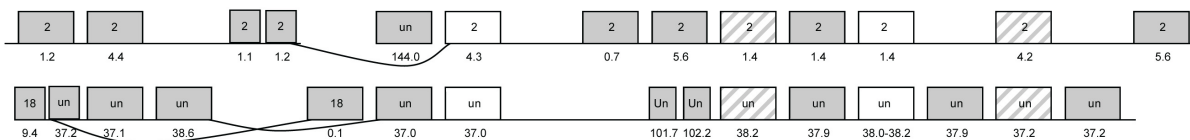


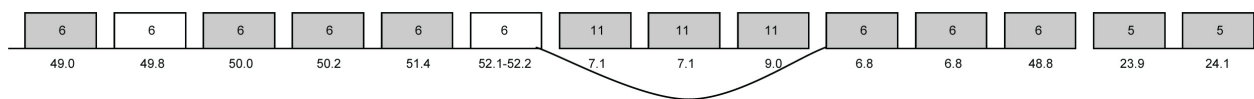
Figure 2

Conserved synteny between human chromosome 17 and chromosomes of other vertebrate species. Picture illustrating conservation of synteny between the species included in this analysis. Gene order is depicted based on the positions in the human genome. Genes located on the same chromosomes or scaffolds in other species are connected with lines in the figure. The names of the gene families are given in the boxes representing the human genes and chromosome number or scaffold number is given in the boxes representing the mouse and fish genes. Numbers below boxes indicate chromosomal or scaffold position in megabases. Two narrow boxes next to each other indicate local duplications. White boxes represent gene families not analyzed with phylogenetic methods in this study (NPY and Hox). Striped boxes indicate genes for which phylogeny is inconclusive but where the positional information is in agreement with the proposed paralogon. Boxes denoted "*" for *Takifugu rubripes* Hox clusters indicates that the Hox cluster is scattered on multiple scaffolds, due to incomplete assembly of the genome. Note that two separate parts of *Tetraodon nigroviridis* chromosome 2 displays conserved synteny with human chromosome 2 and 17.

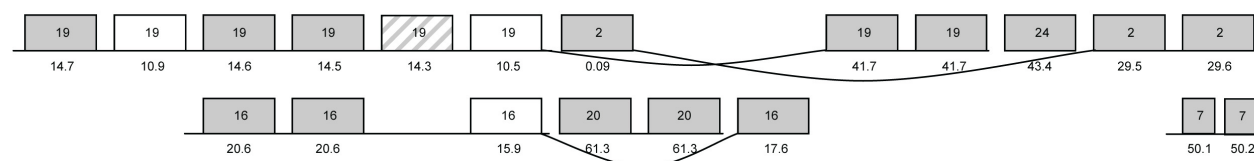
Hsa 7



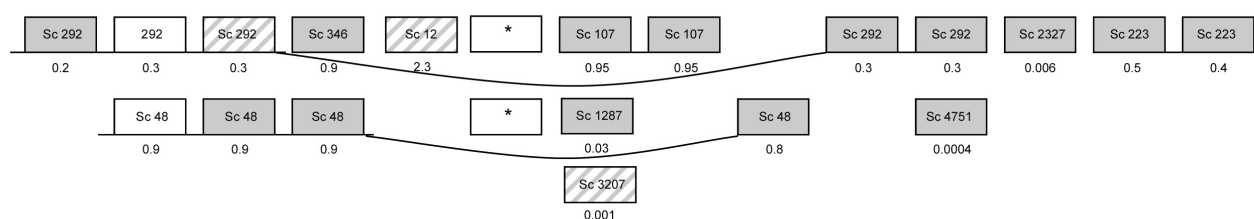
Mmu



Dre



Tru



Tni

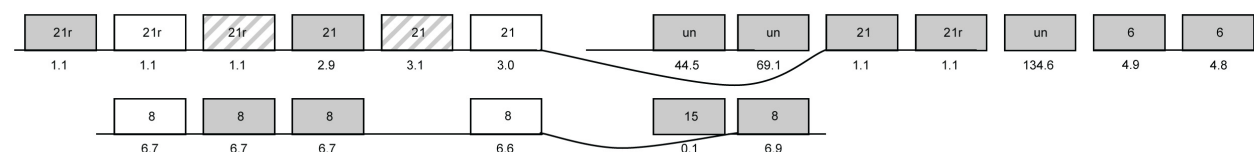


Figure 3
Conserved synteny between human chromosome 7 and chromosomes of other vertebrate species. Gene order, gene name, position, local duplications and linkage of genes are indicated in the same way as in Fig. 2.

ily studied here is comprised by the six members with high-affinity binding to insulin and is represented on all four of the human HOX-bearing chromosomes. On both Hsa 2 and 7 they appear as gene pairs resulting from a duplication before the separation of the branch leading to the mammals and the branch leading to the teleosts (see discussion). Both of these pairs display retained 3R copies in at least one fish species for each gene (Fig. 1A). One *Takifugu rubripes* gene, called Tru.sc149.a in Fig. 1A, appears in an unexpected position in the tree. However, this sequence has a deviating segment compared to all other IGFBP sequences, probably due to difficulties with exon-intron prediction in the database. In our quartet puzzling maximum likelihood analysis (Additional file 2), this gene product clusters with the *Tetraodon nigroviridis* sequence Tni.2, favoring orthology in agreement with the chromosomal position (Fig. 2). This family has

recently been studied by others [54] giving a similar topology but with a different interpretation of the evolutionary history of this gene family (see discussion).

NFE2 (Nuclear factor erythroid 2) family members are involved in gene transcription and due to the presence of four genes located in the vicinity of all four Hox clusters, this family has previously been suggested to have had a similar evolutionary history as the Hox clusters [55,56]. Our results are compatible with such a scenario but the topology in our tree is somewhat difficult to interpret (Fig. 1B), especially regarding identification of fish orthologs. This is probably due to problems with obtaining a reliable alignment for this family because the paralogs only have a short conserved domain in otherwise divergent sequences. Nevertheless, all genes display conserved synteny and have evolved in a time frame consistent with 2R.

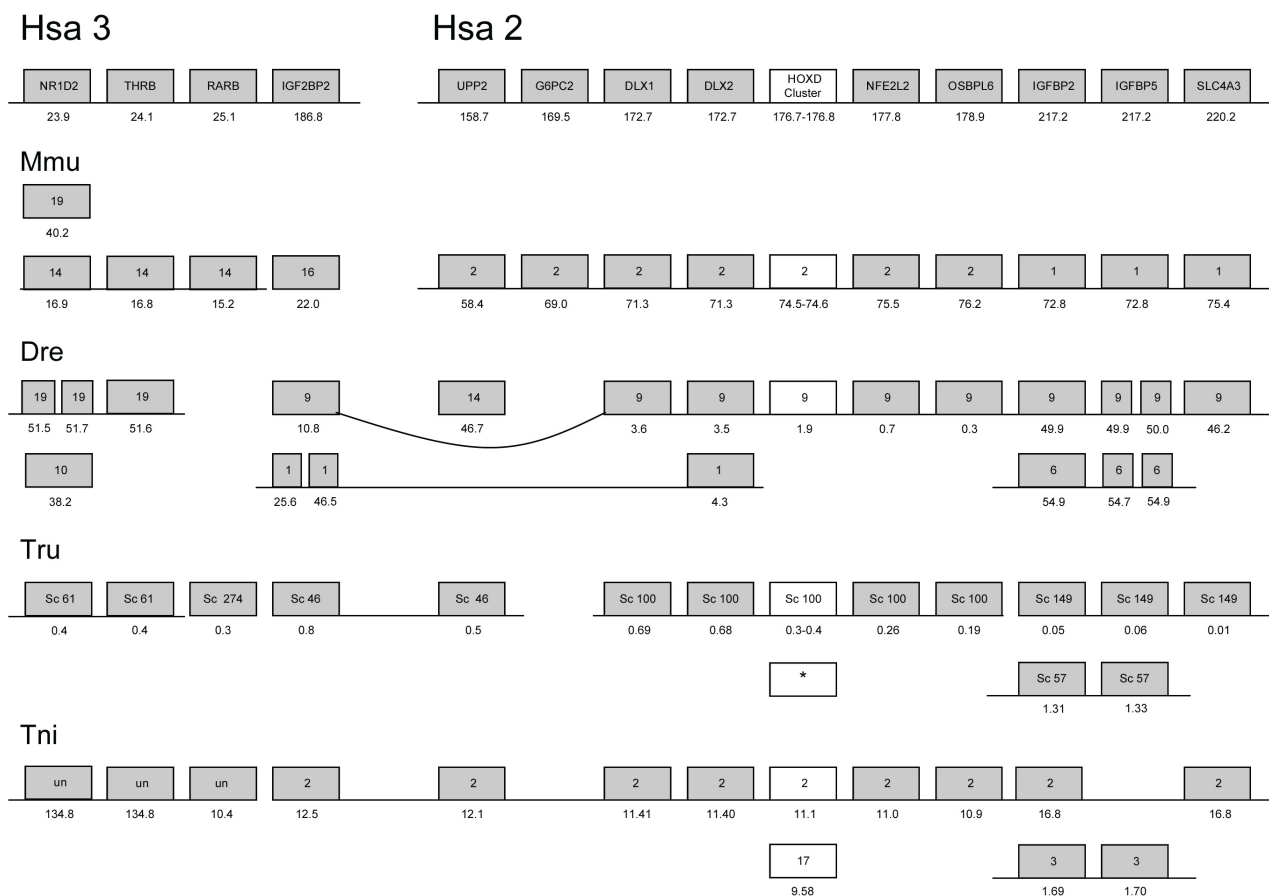


Figure 4
Conserved synteny between human chromosomes 2/3 and chromosomes of other vertebrate species. Gene order, gene name, position, local duplications and linkage of genes are indicated in the same way as in Fig. 2.

Gene families represented on three human Hox chromosomes

DLX genes form gene pairs on three of the chromosomes (2, 7 and 17) in the Hox paralogon. Three of these human genes (one on each chromosome) cluster in the phylogenetic trees with genes that were duplicated in fishes (Additional files, figures 1B and 2B). The *Dlx* family is a widely studied family with important functions in vertebrate development [57]. Our results confirm an early local duplication followed by chromosome duplication for this family as previously suggested [58-61].

IGF2BP, IGF-2-mRNA binding proteins, also called IMPs, are part of an mRNA localization and transport system. The family has members on three of the human chromosomes analyzed in the study. It has previously been suggested that the family originated by duplications before the vertebrate radiation [62,63]. Our phylogenetic analy-

ses support an origin in 2R and also an expansion in the zebrafish in 3R (Additional files, figures 1E and 2E).

SLC4A is a family that is part of the solute carrier superfamily and catalyzes transmembrane bicarbonate transport in the anion exchanger (AE) subfamily [64]. Others have recently analyzed a larger part of this superfamily resulting in a similar tree topology [54]. The family has members on Hsa 2, 7 and 17 and as in many of the other gene families, only one case where both copies resulting from 3R have been retained (Additional files, figures 1K and 2K).

SMARCD, SWI/SNF-related, matrix-associated, actin-dependent regulators of chromatin (SMARC) is a family involved in chromatin remodeling complexes [65]. We have studied one subfamily with three human members (located on chromosomes 7, 12 and 17). It seems like a lineage-specific duplication has occurred in *Ciona*, but the

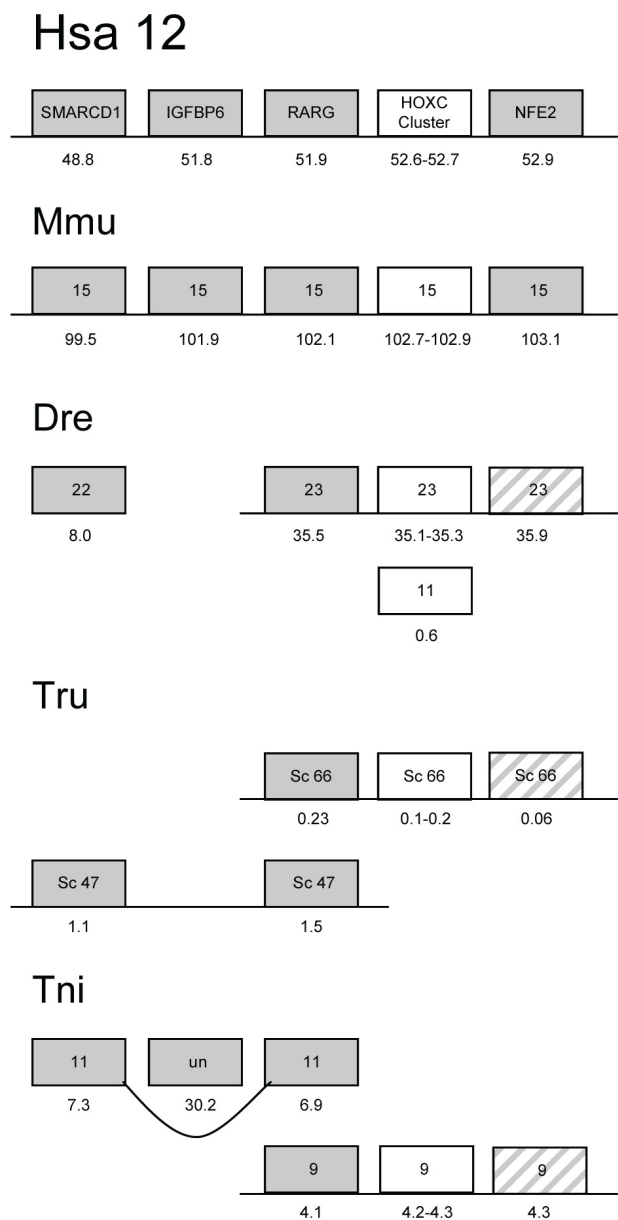


Figure 5
Conserved synteny between human chromosome 12 and chromosomes of other vertebrate species. Gene order, gene name, position, local duplications and linkage of genes are indicated in the same way as in Fig. 2.

rest of the family has expanded in a time window in agreement with 2R (Additional files, figures 1L and 2L).

OSBPL, Oxysterol-binding proteins are a large family involved in diverse cellular processes with an oxysterol binding domain as common feature [66,67] and we have

studied one subfamily with the human members located on chromosomes 2, 7 and 17 (Fig. 6). In one case 3R is clearly visible in the tree (Additional files, figures 1I and 2I).

RAR, Retinoic acid receptors, a gene family belonging to the nuclear receptor superfamily (see below).

Gene families represented on two human *hox* chromosomes

Nuclear receptors

The gene families belonging to the nuclear receptor superfamily, *RAR*, *THR* and *NR1D*, have been studied previously [68,69] with similar results as obtained in this study. The *NR1D* and *THR* families have members on Hsa 3 and 17, and the *RAR* family also includes a member on Hsa 12. Some duplicates from 3R are also retained (Additional files, figures 1H, J and 1M and 2H, J and 2M) and these display conserved synteny with their mammalian orthologs (Fig. 2, 4 and 5). It should be noted that the timing for the duplications is uncertain in the *NR1D* family with one *Branchiostoma floridae* sequence clustering together with orthologs of the vertebrate *NR1D1* sequences, however with low bootstrap support (see Additional files, figures. 1H and 2H).

AOC, (amine oxidase) gene family (Fig. 1C) belongs to the copper-binding amino oxidase superfamily. The human members in this study include one on chromosome 7 and two on chromosome 17. The chromosomal locations are as expected by block duplication in early vertebrate evolution (Fig. 1C and Additional files, figures 1A and 2A).

G6PC (glucose-6-phosphatase beta) (Fig. 1D) is an enzyme involved in both gluconeogenesis and glycogenolysis and has two members on Hsa 17 and one on Hsa 2 [70]. Due to lack of sequences from *Ciona* and *Branchiostoma floridae*, the time window when the expansion of this family took place is wide. The duplication on Hsa 17 is present also in the fishes, indicating origin before the divergence of sarcopterygians and actinopterygians. In addition, the teleost fishes have a local duplication of one of these orthologs resulting in three members of this family located on the same chromosome (Fig. 2). For one of these genes it seems like the zebrafish has retained both of the copies formed in 3R (Fig. 1D and Additional files, figures 1C and 2C). Furthermore, the *Dre3* gene has undergone a recent duplication in the zebrafish (Fig. 1D), although this could also be a result of an assembly error in the database.

MPP, is a subfamily of the membrane-associated guanylate kinase homologs superfamily, an old family involved in cell signaling [71]. Our subfamily has members on Hsa

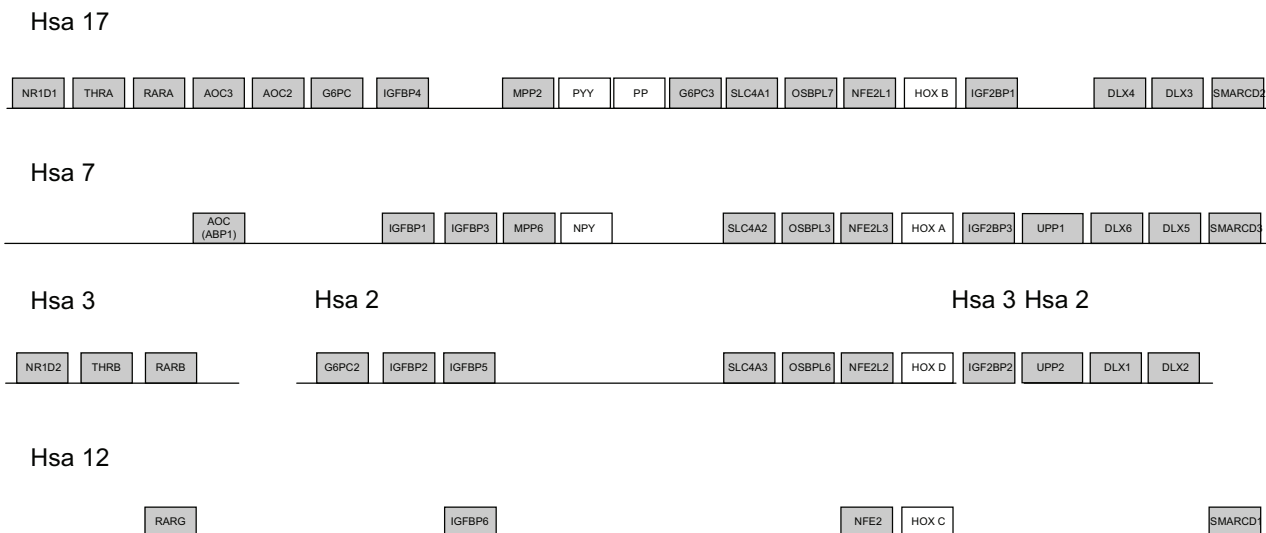


Figure 6
Schematic picture of human paralogs. Schematic picture illustrating the human paralogon analyzed in this study. Gene order as on chromosome 17. Observe that figure is not drawn to scale.

7 and 17, the tree displays a topology consistent with 3R (Additional files, figures 1F and 2F).

UPP, uridine phosphorylate is a small family that catalyzes the reversible phosphorylation of uridine, deoxyuridine and thymidine [72]. It consists of only two human members that seem to have evolved in the same time frame as the proposed 2R events (Additional files, figures 1N and 2N).

Discussion

Extending our previous observation that the two NPY-family genes NPY and PYY are located near the Hox clusters [73], we recently reported that these synteny exist also in the pufferfishes *Takifugu rubripes* and *Tetraodon nigroviridis*, the medaka *Oryzias latipes*, the stickleback *Gasterosteus aculeatus* and the zebrafish *Danio rerio* [74]. Thus, these synteny are likely to have existed in the common ancestor of tetrapods (and other sarcopterygians) and actinopterygians. These chromosome regions have undergone further duplications in the teleost fish tetraploidization, 3R, resulting in duplicates of both the Hox clusters and the NPY and PYY genes. Several other gene families have been suggested to have been duplicated together with the Hox clusters [7-9,54]. To our knowledge the evolutionary histories of only a few families have been investigated in the Hox paralogon using both sequence phylogenies and synteny comparisons, for example the DLX family [58,59], the nuclear receptor family [68] and the voltage-gated sodium channels [75,76]. In the present

study, we have investigated in detail several additional gene families that are represented on multiple chromosomes in the Hox paralogon. With the NPY-family genes as starting points in *Takifugu rubripes*, *Danio rerio* and human, we identified 14 gene families that are present on two, three, or all four of the human Hox chromosomes. We have studied the phylogenies of these gene families initially with the neighbor-joining method and subsequently with quartet puzzling maximum likelihood, and report here that 13 of the 14 families have members that seem to have been duplicated in the early stages of vertebrate evolution (see table 1). The remaining family, G6PC, also has a topology consistent with 2R but due to lack of urochordate or cephalochordate sequences the time window is wider when these expansions could have occurred. The inclusion of human chromosome 3 as part of this paralogon has been suggested earlier [7,75]. However, it is not clear if paralogs on Hsa3 were ancestrally linked to genes on Hsa2 (see figure 4 and [7]) or Hsa 7 [76] before the duplication events. The current analysis is not able to resolve this ambiguity.

Two gene families that have members on all four of the human Hox chromosomes were identified, namely IGFBP and NFE2. The NFE2 family sequences are difficult to align as explained in results and thus the phylogenetic analysis must be interpreted with caution, suffice to conclude that the tree topology (Fig. 1B) is consistent with a quadruplication in early vertebrate evolution. From our analysis it is not possible to deduce if the IGFBP family

Table 1: Summary of the combined interpretation of phylogenetic and positional information

Gene family	2R		3R		
	Phylogenetic support	Comments	Phylogenetic support*	Postional support	Comments
AOC	Yes		No	No	
DLX	Yes		Yes	Yes	Uncertain topology regarding some duplications
G6PC	Yes/No	Uncertain timing, only <i>Drosophila</i> as outgroup	Yes	Yes	Can't exclude lineage specific duplication
IGF2BP	Yes		Yes	Yes	Can't exclude lineage specific duplication
IGFBP	Yes		Yes	Yes	
MPP	Yes		Yes	Yes	Uncertain topology regarding some duplications
NFE	Yes		Yes/No	Yes	Uncertain topology
NR1D	Yes/No	Uncertain topology	No	Yes	Can't exclude lineage specific duplication
OSBPL	Yes		Yes	Yes	Uncertain topology regarding some duplications
RAR	Yes		Yes	Yes	
SLC4A	Yes		Yes	Yes	
SMARCD	Yes		Yes	Yes	Can't exclude lineage specific duplication
THR	Yes		Yes	Yes	
PP	Yes	U	Yes	Yes	Can't exclude lineage specific duplication

* Duplicates in one fish species for one of the mammalian sequences have been considered to be support for 3R.

had one or two members at the time of divergence of the common ancestor of actinopterygians and sarcopterygians from the tunicate lineage (Fig. 1A). The local duplication event could have occurred in early vertebrate evolution, whereupon the gene pair was duplicated as a unit, most likely concomitantly with Hox (i.e. in 2R). Two of the IGFBP duplicates seem to have been lost before the divergence of actinopterygians and sarcopterygians. This scenario is one plausible explanation of the present situation in the human genome with IGFBP gene pairs on Hsa 2 and 7 and single genes on Hsa 12 and 17 (see Fig. 6). The most parsimonious explanation for the observed topology is that loss of genes took place between 1R and 2R. Abbasi et al. [54] interpreted this tree differently, stating that the most parsimonious explanation is two genome duplications followed by two local duplications on chromosome 2 and 7 respectively. In our opinion it is not possible to interpret the tree topology this way. If the local duplications occurred after the whole genome duplications one would expect a different topology, with the locally duplicated genes on Hsa7 clustering together and the ones on Hsa2 clustering together. This is not the case (see Fig. 1A). Their alternative hypothesis includes three whole genome duplications and two losses. In order for this scenario to be compatible with the topology and localization two translocations must have occurred. This involves two additional steps and it also suggests that three whole genome duplications have occurred, but this is not observed for any of the other families in their study.

Six gene families were found to be represented on three of the human Hox chromosomes, i.e., DLX, IGF2BP, SLC4A, SMARCD, OSBPL and RAR. Note that the DLX family underwent a local duplication before the chromosome duplications, in analogy with the IGFBP family described above, as shown by the phylogenetic tree in Fig. 1C and the chromosome maps in Figs. 2, 3, 4, 5, 6. This family has previously been difficult to resolve using sequence analyses in mammals, but thanks to the many fish sequences it can be established that DLX1, 4 and 6 form one series of paralogs and that DLX 2, 3 and 5 form another [58].

Representation on two of the mammalian Hox chromosomes was found for six gene families, namely two of the three nuclear receptor families THR and NR1D, as well as AOC, G6PC, MPP, and UPP. The AOC family (Fig. 1C), for instance, is represented on Hsa7 and Hsa17 with orthologs located in the corresponding regions in the mouse genome. A local duplication has occurred in mammals, resulting in AOC2 and AOC3 on Hsa17, with loss of AOC2 in mouse. There is also a third member located on chromosome 17 – a pseudogene that was not present in version 43 of the Ensembl database [77]. A series of duplications has occurred in the mouse genome resulting in four copies on Mmu6 [78]. In contrast, no additional duplications have occurred among the fish species, not even in 3R.

The G6PC family (Fig. 1D) is also represented on two chromosomes in human (Figs. 2 and 4). The local duplication observed on Hsa17 is present also in the three teleosts studied, showing that it happened before the sarcopterygian-actinopterygian split (Fig. 2). Interestingly, the three teleosts have an additional local duplication of one of these genes, not present in the mammals. In this case, zebrafish seems to have retained a 3R duplicate of one of these genes (Dre12a) which has a copy on Dre3.

These gene families show that local duplications occur frequently and thereby complicate delineation of gene family histories that involve both local and chromosome duplications. To distinguish between locally duplicated genes and gene duplicates resulting from large-scale events such as 2R, the use of several species is crucial. It has recently been suggested that gene families located on the Hox-bearing chromosomes were not duplicated together with Hox based on the observed differences in the topologies [54]. Instead the gene families were proposed to have duplicated in several independent small-scale events and later translocated to the same chromosomes [54]. This suggests independent translocation events before the divergence of actinopterygians and sarcopterygians. In the light of the results from large-scale studies describing duplicated regions in the genomes of extant vertebrates and the reconstruction of ancestral chromosomes [10,11] it seems unlikely that the observed pattern arose by many independent duplications followed by translocations to the same chromosomes. Furthermore, the strict demand for (A, B)(C, D)-topologies as the only criteria to support 2R is too restrictive due to complications such as unequal evolutionary rates, gene conversion and loss of genes after duplication. Thus, trees not showing the perfect (A, B)(C, D)-topology are still compatible with large-scale duplication events if the relative dating of these events agrees with positional information from several species.

It was expected that several of the gene families would display fewer than four members in mammals, and likewise fewer than eight members in the teleosts (Figs. 2, 3, 4, 5). As has been described on numerous previous occasions, gene losses are common after duplications why at least one member is often missing. Striking examples of gene loss are seen in the Hox clusters, where the ancestral set of 14 genes should have resulted in 56 genes after the block quadruplication. However, one of the ancestral Hox genes (Hox14) has lost all four paralogs in mammals after the duplications [79-81], and several other Hox genes have lost 1-3 of the duplicates, resulting in the present number of 39 Hox genes in mammalian genomes. Furthermore, both the pufferfishes and zebrafish seem to have lost a whole Hox cluster, albeit different ones in these two lineages [12]. As shown in Fig. 6, information from families

with only two members still gives important information when they are interpreted together to define regions that have been duplicated. If loss of genes from duplicated chromosomal segments occurs independently, positional information from families with two members taken together still gives valuable information.

In some of the gene families the teleost-specific duplicates are only represented in one species (Figs. 2, 3, 4, 5). This made it difficult to distinguish lineage-specific duplications from 3R duplications followed by loss of the duplicate in some lineages. We believe the poor representation of 3R duplicates in our trees is partly a reflection of the strict criteria we used when editing the alignments of sequences from incomplete or poorly annotated genome databases. Because sequences lacking the analyzed domains were excluded, our dataset probably underestimates the real number of family members in the included fish species. Some of the genes lacking domains could be on their way to become pseudogenes, but there could also be problems in the prediction of coding regions in the fish genomes, for example due to short introns in *Tetraodon nigroviridis* [25]. Even though we started our analysis in the fish genomes and therefore should not be biased towards any of the human chromosomes, one of them stands out as more poorly represented than the other three with regard to the number of gene families, namely human chromosome 12. Uneven retention of gene duplicates on different chromosomes resulting from a quadruplication has been observed previously for the paralogon comprised by the human chromosomes 4, 5, 8 and 10 [82] as well as for chromosomes 1, 6, 9 and 19 [83]. This could be the result of more extensive rearrangements on some chromosomes, causing loss of genes for instance by affecting gene regulatory elements.

In studies of gene duplications the dating of the duplication events is of course an important factor. This can be done using molecular clocks [84], ideally calibrated against known speciation events. Dating of duplication events can also be obtained by using information from several species, thereby making it possible to relatively date the origin of genes in relation to speciation events. The 2R tetraploidizations have been proposed to have occurred after the split of urochordates and vertebrates but before the origin of gnathostomes [9]. We used sequences from the tunicates *Ciona intestinalis* or *Ciona savignyi* or the cephalochordate *Branchiostoma floridae* as outgroup to see if the duplications we investigated took place in the vertebrate lineage, as the question we wanted to address was whether the syntenic gene families expanded in the same time window. So far it has not been possible to determine unambiguously whether both of the 2R tetraploidizations took place before the origin of cyclostomes, or if the second tetraploidization occurred

on the gnathostome branch after it had separated from the cyclostomes mainly due to lack of sequence information.

Many types of genetic events complicate efforts to reconstruct the evolutionary history of individual gene families, not only frequent duplications and losses, but also differences in the rate of change between members in a gene family or in the same gene over time. Another source of difficulties that we have experienced, as mentioned above, is uncertainties in protein alignments due to duplication of domains in some family members. Because sequence-based phylogenetic analyses and chromosome positions (synteny) constitute two quite different types of information, they can be combined to deduce evolutionary schemes in a more reliable way. The combined use of synteny and phylogenetic information makes it possible to delineate old and new duplications, identify translocations of genes and also allows investigation of gene families with only two members. Mechanisms like crossing-over and gene conversion may still obscure relationships, but at least the latter is expected to decrease with time as the family members accumulate differences that reduce the likelihood of such events. Particularly gene families with very short sequences, and/or highly conserved or rapidly diverging sequences, can be resolved by considering chromosomal position, as has been shown for the insulin/relaxin family of peptides [85], the NPY family of peptides as mentioned above [74], and some of the NPY-family receptors [86] as well as for example the NFE2 family in this study. All of these families were found to have expanded in the early vertebrate tetraploidizations.

Conclusion

In conclusion, our study shows that the previously well-studied Hox cluster was certainly not the only gene family in this chromosomal segment that was quadrupled in early vertebrate evolution: the duplicated regions extend quite far in both directions from the Hox clusters and involve a larger number of gene families that are totally unrelated to the Hox-gene sequences. Many genes have been lost in these families after the quadruplication, like in the Hox clusters, thereby obscuring the duplication events. Nevertheless, the overall pattern of genes in the regions flanking the four Hox clusters clearly shows that the quadruplication encompasses a major chromosomal segment, probably a complete ancestral chromosome.

Methods

Identification of chromosomal regions

The selection of the 14 gene families used in this study was based on the location of family members close to the genes coding for NPY family peptides in fugu, *Takifugu rubripes* and zebrafish, *Danio rerio*. The gene families included have genes located on at least two of the human

chromosomes 2, 3, 7, 12 and 17. Chromosome 3 is a non-Hox-bearing chromosome but earlier studies have suggested that it is a part of the Hox-paralogue [7,8,75]. The two gene families IGFBP and SLC4A were included because they were already known to have copies on several of the chromosomes in the paralogue and were located near the human NPY-family genes.

Database searches

The protein family predictions in Ensembl database version 43 <http://www.ensembl.org/> were used together with relevant literature and BLAST [87] to identify additional sequences not included in the Ensembl protein families. All amino acid sequences representing the longest transcripts of every member of the 14 gene families in *Tetraodon nigroviridis*, *Takifugu rubripes*, *Danio rerio*, *Mus musculus* and *Homo sapiens* were obtained in order to obtain information of the repertoires in the common ancestor of actinopterygians and sarcopterygians using the Ensembl database version 43. This allows for the identification of shared and lineage specific in teleosts and tetrapods, respectively. For a full set of Ensembl IDs see Additional file 3.

Alignments and phylogenetic analyses

Protein domains in each sequence were identified by searches against the Pfam database <http://www.sanger.ac.uk/Software/Pfam/> and aligned using the Windows version of Clustal × 1.81 [88,89]. Sequences lacking described domains, or incomplete domains, were removed from the alignment (for description of which criteria that were used for each family see figure legends in supplementary file 1 and 2). Alignments were thereafter manually inspected and edited to remove poorly aligned sequences. Phylogenetic trees were constructed using the neighbor-joining (NJ) method with standard settings (Gonnet weight matrix, gap opening penalty 10.0 and gap extension penalty 0.20) as implemented in the windows version of Clustal W 1.81 [88] with 1000 bootstrap replicates. Bootstrap values below 50% were considered non-supportive. Sequences from *Drosophila melanogaster* and *Ciona intestinalis* (or *Ciona savignyi* or *Branchiostoma floridae*) were used as outgroups in the phylogenetic analyses in order to relatively date duplications. Quartet puzzling maximum-likelihood trees were constructed using Tree-Puzzle 5.2 [90] (for Tree-puzzle settings in each analysis, see supplementary file 2).

Chromosomal maps

Based on the phylogenetic analyses, positional information from gene family members duplicated after the split of urochordates and the rest of the chordates and before the split of sarcopterygians and actinopterygians was used to draw chromosomal maps. This allowed for identifica-

tion gene families that most likely were syntenic in the ancestor of all vertebrates.

Authors' contributions

GS performed the database searches, phylogenetic and chromosome analyses and wrote part of the manuscript. TAL participated in phylogenetic and chromosome analyses and writing of the manuscript. DL initiated the study and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Additional material

Additional file 1

Neighbor-joining trees for the 14 gene families analyzed.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-254-S1.pdf>]

Additional file 2

Quartet puzzling trees for the 14 gene families analyzed.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-254-S2.pdf>]

Additional file 3

Accession numbers, positional information and description of sequences used in the phylogenetic analyses.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-8-254-S3.xls>]

Acknowledgements

The authors wish to thank Susanne Dreborg for help with preparing figures and critically reading the manuscript. This work was supported by grants from the Swedish Research Council and Carl Trygger's Foundation. The authors would like to thank the anonymous reviewers for comments improving the manuscript.

References

- Coulier F, Popovici C, Villet R, Birnbaum D: **MetaHox gene clusters.** *J Exp Zool* 2000, **288(4)**:345-351.
- Lundin LG: **Gene homologies with emphasis on paralogous genes and chromosomal regions.** *Life Sci Adv (Genet)* 1989, **(8)**:89-104.
- Popovici C, Leveugle M, Birnbaum D, Coulier F: **Homeobox gene clusters and the human paralogy map.** *FEBS Letters* 2001, **491(3)**:237-242.
- Popovici C, Leveugle M, Birnbaum D, Coulier F: **Coparalogy: physical and functional clusterings in the human genome.** *Biochem Biophys Res Commun* 2001, **288(2)**:362-370.
- Pebusque M, Coulier F, Birnbaum D, Pontarotti P: **Ancient large-scale genome duplications: phylogenetic and linkage analyses shed light on chordate genome evolution.** *Mol Biol Evol* 1998, **15(9)**:1145-1159.
- Lundin LG: **Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse.** *Genomics* 1993, **16(1)**:1-19.
- Lundin LG, Larhammar D, Hallbook F: **Numerous groups of chromosomal regional paralogies strongly indicate two genome doublings at the root of the vertebrates.** *J Struct Funct Genomics* 2003, **3(1-4)**:53-63.
- Larhammar D, Lundin LG, Hallbook F: **The human Hox-bearing chromosome regions did arise by block or chromosome (or even genome) duplications.** *Genome Research* 2002, **12(12)**:1910-1920.
- Panopoulou G, Poustka AJ: **Timing and mechanism of ancient vertebrate genome duplications – the adventure of a hypothesis.** *Trends Genet* 2005, **21(10)**:559-567.
- Nakatani Y, Takeda H, Kohara Y, Morishita S: **Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates.** *Genome Research* 2007, **17(9)**:1254-1265.
- Dehal P, Boore JL: **Two rounds of whole genome duplication in the ancestral vertebrate.** *PLoS Biology* 2005, **3(10)**:e314.
- Hoegg S, Meyer A: **Hox clusters as models for vertebrate genome evolution.** *Trends Genet* 2005, **21(8)**:421-424.
- Garcia-Fernandez J, Holland PV: **Archetypal organization of the amphioxus Hox gene cluster.** *Nature* 1994, **370(6490)**:563-566.
- Venkatesh B, Kirkness EF, Loh YH, Halpern AL, Lee AP, Johnson J, Dandona N, Viswanathan LD, Tay A, Venter JC, et al.: **Survey Sequencing and Comparative Analysis of the Elephant Shark (*Callorhynchus milii*) Genome.** *PLoS Biology* 2007, **5(4)**:e101.
- Monteiro AS, Ferrier DE: **Hox genes are not always Colinear.** *Int J Biol Sci* 2006, **2(3)**:95-103.
- Spagnuolo A, Ristoratore F, Di Gregorio A, Aniello F, Branno M, Di Lauro R: **Unusual number and genomic organization of Hox genes in the tunicate *Ciona intestinalis*.** *Gene* 2003, **309(2)**:71-79.
- Seo HC, Edvardsen RB, Maeland AD, Bjordal M, Jensen MF, Hansen A, Flaot M, Weissenbach J, Lehrach H, Wincker P, et al.: **Hox cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*.** *Nature* 2004, **431(7004)**:67-71.
- Ikuta T, Saiga H: **Organization of Hox genes in ascidians: present, past, and future.** *Dev Dyn* 2005, **233(2)**:382-389.
- Ferrier DE, Holland PV: **Ancient origin of the Hox gene cluster.** *Nat Rev Genet* 2001, **2(1)**:33-38.
- Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL, et al.: **Zebrafish hox clusters and vertebrate genome evolution.** *Science* 1998, **282(5394)**:1711-1714.
- Naruse K, Fukamachi S, Mitani H, Kondo M, Matsuoka T, Kondo S, Hanamura N, Morita Y, Hasegawa K, Nishigaki R, et al.: **A detailed linkage map of medaka, *Oryzias latipes*: comparative genomics and genome evolution.** *Genetics* 2000, **154(4)**:1773-1784.
- Crow KD, Stadler PF, Lynch VJ, Amemiya C, Wagner GP: **The "fish-specific" Hox cluster duplication is coincident with the origin of teleosts.** *Mol Biol Evol* 2006, **23(1)**:121-136.
- Taylor JS, Braasch I, Frickey T, Meyer A, Peer Y Van de: **Genome duplication, a trait shared by 22000 species of ray-finned fish.** *Genome Research* 2003, **13(3)**:382-390.
- Taylor JS, Peer Y Van de, Braasch I, Meyer A: **Comparative genomics provides evidence for an ancient genome duplication event in fish.** *Philos Trans R Soc Lond B Biol Sci* 2001, **356(1414)**:1661-1679.
- Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, et al.: **Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype.** *Nature* 2004, **431(7011)**:946-957.
- Christoffels A, Koh EG, Chia JM, Brenner S, Aparicio S, Venkatesh B: **Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes.** *Mol Biol Evol* 2004, **21(6)**:1146-1151.
- Vandepoele K, De Vos W, Taylor JS, Meyer A, Peer Y Van de: **Major events in the genome evolution of vertebrates: paraneome age and size differ considerably between ray-finned fishes and land vertebrates.** *Proc Natl Acad Sci USA* 2004, **101(6)**:1638-1643.
- Hoegg S, Brinkmann H, Taylor JS, Meyer A: **Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish.** *J Mol Evol* 2004, **59(2)**:190-203.
- Venkatesh B: **Evolution and diversity of fish genomes.** *Curr Opin Genet Dev* 2003, **13(6)**:588-592.

30. Meyer A, Schartl M: **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* 1999, **11(6)**:699-704.
31. Crow KD, Wagner GP: **What is the role of genome duplication in the evolution of complexity and diversity?** *Mol Biol Evol* 2006, **23(5)**:887-892.
32. Hurley IA, Mueller RL, Dunn KA, Schmidt EJ, Friedman M, Ho RK, Prince VE, Yang Z, Thomas MG, Coates MI: **A new time-scale for ray-finned fish evolution.** *Proc Biol Sci* 2007, **274(1609)**:489-498.
33. Larhammar D, Risinger C: **Molecular genetic aspects of tetraploidy in the common carp *Cyprinus carpio*.** *Mol Phylogenet Evol* 1994, **3(1)**:59-68.
34. Allendorf FW, Thorgaard GH: **Tetraploidy and the Evolution of Salmonid Fishes.** In *The Evolutionary Genetics of Fishes* Edited by: Turner BJ. Plenum Press; 1984:1-53.
35. Uyeno T, Smith GR: **Tetraploid origin of the karyotype of catostomid fishes.** *Science* 1972, **175(22)**:644-646.
36. David L, Blum S, Feldman MV, Lavi U, Hillel J: **Recent duplication of the common carp (*Cyprinus carpio* L.) genome as revealed by analyses of microsatellite loci.** *Mol Biol Evol* 2003, **20(9)**:1425-1434.
37. Le Comber SC, Smith C: **Polyploidy in fishes: patterns and processes.** *Biological Journal of the Linnean Society* 2004, **82**:431-442.
38. Lynch M, Conery JS: **The evolutionary fate and consequences of duplicate genes.** *Science* 2000, **290(5494)**:1151-1155.
39. Amoree A, Suzuki T, Yan YL, Pomeroy J, Singer A, Amemiya C, Postlethwait JH: **Developmental roles of pufferfish *Hox* clusters and genome evolution in ray-fin fish.** *Genome Research* 2004, **14(1)**:1-10.
40. Zhang J: **Evolution by gene duplication: an update.** *Trends Ecol Evol* 2003, **18(6)**:292-298.
41. Larhammar D, Salaneck E: **Molecular evolution of NPY receptor subtypes.** *Neuropeptides* 2004, **38(4)**:141-151.
42. Kellis M, Birren BW, Lander ES: **Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*.** *Nature* 2004, **428(6983)**:617-624.
43. Scannell DR, Byrne KP, Gordon JL, Wong S, Wolfe KH: **Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts.** *Nature* 2006, **440(7082)**:341-345.
44. Seoighe C, Wolfe KH: **Extent of genomic rearrangement after genome duplication in yeast.** *Proc Natl Acad Sci USA* 1998, **95(8)**:4447-4452.
45. Woods IG, Wilson C, Friedlander B, Chang P, Reyes DK, Nix R, Kelly PD, Chu F, Postlethwait JH, Talbot WS: **The zebrafish gene map defines ancestral vertebrate chromosomes.** *Genome Research* 2005, **15(9)**:1307-1314.
46. Brunet FG, Crollius HR, Paris M, Aury JM, Gibert P, Jaillon O, Laudet V, Robinson-Rechavi M: **Gene loss and evolutionary rates following whole-genome duplication in teleost fishes.** *Mol Biol Evol* 2006, **23(9)**:1808-1816.
47. Bowers JE, Chapman BA, Rong J, Paterson AH: **Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events.** *Nature* 2003, **422(6930)**:433-438.
48. Seoighe C, Gehring C: **Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome.** *Trends Genet* 2004, **20(10)**:461-464.
49. Wolfe KH: **Yesterday's polyploids and the mystery of diploidization.** *Nat Rev Genet* 2001, **2(5)**:333-341.
50. Simillion C, Vandepoele K, Van Montagu MC, Zabeau M, Peer Y Van de: **The hidden duplication past of *Arabidopsis thaliana*.** *Proc Natl Acad Sci USA* 2002, **99(21)**:13627-13632.
51. Vandepoele K, Simillion C, Peer Y Van de: **Detecting the undetectable: uncovering duplicated segments in *Arabidopsis thaliana* by comparison with rice.** *Trends Genet* 2002, **18(12)**:606-608.
52. Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, et al.: **The amphioxus genome and the evolution of the chordate karyotype.** *Nature* 2008, **453(7198)**:1064-1071.
53. Hwa V, Oh Y, Rosenfeld RG: **The insulin-like growth factor-binding protein (IGFBP) superfamily.** *Endocr Rev* 1999, **20(6)**:761-787.
54. Abbasi AA, Grzeschik KH: **An insight into the phylogenetic history of HOX linked gene families in vertebrates.** *BMC Evol Biol* 2007, **7(1)**:239.
55. Pratt SJ, Drejer A, Foot H, Barut B, Brownlie A, Postlethwait J, Kato Y, Yamamoto M, Zon LI: **Isolation and characterization of zebrafish NFE2.** *Physiol Genomics* 2002, **11(2)**:91-98.
56. Kobayashi A, Ito E, Toki T, Kogame K, Takahashi S, Igarashi K, Hayashi N, Yamamoto M: **Molecular cloning and functional characterization of a new Cap'n' collar family transcription factor Nrf3.** *J Biol Chem* 1999, **274(10)**:6443-6452.
57. Panganiban G, Rubenstein JL: **Developmental functions of the Distal-less/Dlx homeobox genes.** *Development* 2002, **129(19)**:4371-4386.
58. Stock DW: **The Dlx gene complement of the leopard shark, *Triakis semifasciata*, resembles that of mammals: implications for genomic and morphological evolution of jawed vertebrates.** *Genetics* 2005, **169(2)**:807-817.
59. Stock DW, Ellies DL, Zhao Z, Ekker M, Ruddle FH, Weiss KM: **The evolution of the vertebrate Dlx gene family.** *Proc Natl Acad Sci USA* 1996, **93(20)**:10858-10863.
60. Sumiyama K, Irvine SQ, Ruddle FH: **The role of gene duplication in the evolution and function of the vertebrate Dlx/distal-less bigene clusters.** *J Struct Funct Genomics* 2003, **3(1-4)**:151-159.
61. Neidert AH, Virupannavar V, Hooker GW, Langeland JA: **Lamprey Dlx genes and early vertebrate evolution.** *Proc Natl Acad Sci USA* 2001, **98(4)**:1665-1670.
62. Nielsen FC, Nielsen J, Kristensen MA, Koch G, Christiansen J: **Cytoplasmic trafficking of IGF-II mRNA-binding protein by conserved KH domains.** *J Cell Sci* 2002, **115(Pt 10)**:2087-2097.
63. Nielsen J, Cilius Nielsen F, Kragh Jakobsen R, Christiansen J: **The biphasic expression of IMP/Vg1-RBP is conserved between vertebrates and *Drosophila*.** *Mech Dev* 2000, **96(1)**:129-132.
64. Sterling D, Casey JR: **Bicarbonate transport proteins.** *Biochem Cell Biol* 2002, **80(5)**:483-497.
65. Ring HZ, Vameghi-Meyers V, Wang W, Crabtree GR, Francke U: **Five SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin (SMARC) genes are dispersed in the human genome.** *Genomics* 1998, **51(1)**:140-143.
66. Lehto M, Olkkonen VM: **The OSBP-related proteins: a novel protein family involved in vesicle transport, cellular lipid metabolism, and cell signalling.** *Biochim Biophys Acta* 2003, **1631(1)**:1-11.
67. Jaworski CJ, Moreira E, Li A, Lee R, Rodriguez IR: **A family of 12 human genes containing oxysterol-binding domains.** *Genomics* 2001, **78(3)**:185-196.
68. Escriba Garcia H, Laudet V, Robinson-Rechavi M: **Nuclear receptors are markers of animal genome evolution.** *J Struct Funct Genomics* 2003, **3(1-4)**:177-184.
69. Laudet V: **Evolution of the nuclear receptor superfamily: early diversification from an ancestral orphan receptor.** *J Mol Endocrinol* 1997, **19(3)**:207-226.
70. Guionie O, Moallic C, Niamke S, Placier G, Sine JP, Colas B: **Identification and primary characterization of specific proteases in the digestive juice of *Archachatina ventricosa*.** *Comp Biochem Physiol B Biochem Mol Biol* 2003, **135(3)**:503-510.
71. Katoh M, Katoh M: **Identification and characterization of human MPP7 gene and mouse Mpp7 gene in silico.** *Int J Mol Med* 2004, **13(2)**:333-338.
72. Johansson M: **Identification of a novel human uridine phosphorylase.** *Biochem Biophys Res Commun* 2003, **307(1)**:41-46.
73. Soderberg C, Wraith A, Ringvall M, Yan YL, Postlethwait JH, Brodin L, Larhammar D: **Zebrafish genes for neuropeptide Y and peptide YY reveal origin by chromosome duplication from an ancestral gene linked to the homeobox cluster.** *J Neurochem* 2000, **75(3)**:908-918.
74. Sundström G, Larsson TA, Brenner S, Venkatesh B, Larhammar D: **Evolution of the neuropeptide Y family: New genes by chromosome duplications in early vertebrates and in teleost fishes.** *Gen Comp Endocrinol* 2008, **155(3)**:705-716.
75. Plummer NW, Meisler MH: **Evolution and diversity of mammalian sodium channel genes.** *Genomics* 1999, **57(2)**:323-331.
76. Novak AE, Jost MC, Lu Y, Taylor AD, Zakon HH, Ribera AB: **Gene duplications and evolution of vertebrate voltage-gated sodium channels.** *J Mol Evol* 2006, **63(2)**:208-221.
77. Schwelberger HG: **The origin of mammalian plasma amine oxidases.** *J Neural Transm* 2007, **114(6)**:757-762.
78. Lundwall A, Malm J, Clauss A, Valtonen-Andre C, Olsson AY: **Molecular cloning of complementary DNA encoding mouse seminal vesicle-secreted protein SVS I and demonstration of**

- homology with copper amine oxidases. *Biol Reprod* 2003, **69(6)**:1923-1930.
79. Minguillon C, Gardenyes J, Serra E, Castro LF, Hill-Force A, Holland PV, Amemiya CT, Garcia-Fernandez J: **No more than 14: the end of the amphioxus Hox cluster.** *Int J Biol Sci* 2005, **1(1)**:19-23.
 80. Powers TP, Amemiya CT: **Evidence for a Hox14 paralog group in vertebrates.** *Curr Biol* 2004, **14(5)**:R183-184.
 81. Ferrier DE: **Hox genes: Did the vertebrate ancestor have a Hox14?** *Curr Biol* 2004, **14(5)**:R210-211.
 82. Vienne A, Rasmussen J, Abi-Rached L, Pontarotti P, Gilles A: **Systematic phylogenomic evidence of en bloc duplication of the ancestral 8p11.21-8p21.3-like region.** *Mol Biol Evol* 2003, **20(8)**:1290-1298.
 83. Abi-Rached L, Gilles A, Shiina T, Pontarotti P, Inoko H: **Evidence of en bloc duplication in vertebrate genomes.** *Nature Genetics* 2002, **31(1)**:100-105.
 84. Durand D, Hoberman R: **Diagnosing duplications – can it be done?** *Trends Genet* 2006, **22(3)**:156-164.
 85. Olinski RP, Lundin LG, Hallbook F: **Conserved synteny between the Ciona genome and human paralogs identifies large duplication events in the molecular evolution of the insulin-relaxin gene family.** *Mol Biol Evol* 2006, **23(1)**:10-22.
 86. Larsson TA, Olsson F, Sundstrom G, Brenner S, Venkatesh B, Larhammar D: **Pufferfish and Zebrafish Have Five Distinct NPY Receptor Subtypes, but Have Lost Appetite Receptors Y1 and Y5.** *Ann NY Acad Sci* 2005, **1040**:375-377.
 87. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215(3)**:403-410.
 88. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic acids research* 1994, **22(22)**:4673-4680.
 89. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ: **Multiple sequence alignment with Clustal X.** *Trends Biochem Sci* 1998, **23(10)**:403-405.
 90. Schmidt HA, Strimmer K, Vingron M, von Haeseler A: **TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing.** *Bioinformatics* 2002, **18(3)**:502-504.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

