

---

## ParaHox Cluster Evolution — Hagfish and Beyond

Author(s): Rebecca F. Furlong and John F. Mulley

Source: Zoological Science, 25(10):955-959. 2008.

Published By: Zoological Society of Japan

DOI: <http://dx.doi.org/10.2108/zsj.25.955>

URL: <http://www.bioone.org/doi/full/10.2108/zsj.25.955>

---

BioOne ([www.bioone.org](http://www.bioone.org)) is a nonprofit, online aggregation of core research in the biological, ecological, and environmental sciences. BioOne provides a sustainable online platform for over 170 journals and books published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Web site, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at [www.bioone.org/page/terms\\_of\\_use](http://www.bioone.org/page/terms_of_use).

Usage of BioOne content is strictly limited to personal, educational, and non-commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

# ParaHox Cluster Evolution — Hagfish and Beyond

Rebecca F. Furlong\* and John F. Mulley

Department of Zoology, Oxford University, South Parks Road, Oxford OX1 3PS, UK

**The ParaHox genes comprise three Hox-related homeobox gene families, found throughout the animals. They were first discovered in the invertebrate chordate amphioxus, where they are tightly clustered. In this paper we carry out a comparative review of ParaHox gene cluster organization among the deuterostomes, and discuss how the recently published hagfish ParaHox clusters fit into current theories about the evolution of this group of genes.**

**Key words:** hagfish, cyclostome, ParaHox, homeobox cluster

## INTRODUCTION TO THE PARAHOX GENE CLUSTER

The ParaHox gene cluster was first discovered in the cephalochordate amphioxus (Brooke et al., 1998), and is composed of members of three Hox-related homeobox gene families: Gsx, Xlox, and Cdx. Gsx is most similar to the anterior Hox genes (insect *lab*, *pb*), Xlox to group 3 Hox genes (insect *zen*), and Cdx to the posterior Hox genes (insect *Abd-B*). Only one member of each ParaHox gene family is found in amphioxus, and these are located in a single cluster with Gsx adjacent to Xlox in the same orientation, followed by Cdx on the opposite strand, the cluster spanning about 56kb (Ferrier et al., 2005) (Table 1). There is no doubt that the clustered organization of these genes as seen in amphioxus is ancestral; the common origin of the Hox and ParaHox gene clusters from a hypothetical ancestral “protoHox” cluster has been discussed extensively elsewhere (e.g., Brooke et al., 1998; Holland, 2001; Garcia-Fernandez, 2005; Chourrout et al., 2006) and will not be discussed in detail here. However, the protoHox hypothesis has also led to other parallels being drawn between the organization and control of the Hox and ParaHox clusters.

The Hox genes are recognised for their spatial and temporal colinearity; that is, the order of expression both during development and along the body axis echoes the order of genes within the cluster. In amphioxus, Gsx is expressed in the cerebral vesicle, Xlox is expressed transiently in the neural tube and in the presumptive gut, and Cdx is expressed in the posterior neural tube and gut. Thus the order of expression sites along the body axis echoes the order of ParaHox genes within the cluster and may be described as spatial colinearity (Brooke et al., 1998). It has been suggested that this expression pattern is modified from an ancestral spatial colinearity of gut expression present at the genesis of the ParaHox cluster (Holland, 2001). Furthermore there is evidence of reverse temporal colinearity, with Cdx expressed first, followed by Xlox and finally Gsx. But with only three genes in a cluster, there are only six possible

orders of expression, two of which will be co-linear with the gene order. Could the colinearity simply have occurred by chance?

The cephalochordates (amphioxus) were long thought to be the closest invertebrate relatives of the vertebrates; however, recent studies have revealed that the urochordates form a sister group to the vertebrates (Delsuc et al., 2006; Bourlat et al., 2006; Vienne and Pontarotti, 2006). Amphioxus remains a useful model for its apparently archetypical genomic organization and gene complement (Garcia-Fernandez and Holland, 1994), but the genomes of other deuterostome animals may also reveal much about the evolution of ParaHox gene clustering and expression. For example, in the invertebrate urochordate *Ciona intestinalis*, surveys of the ParaHox genes indicated that their organization was at the very least much more dispersed than that in amphioxus, and that they may not be chromosomally linked (Ferrier and Holland, 2002; Wada et al., 2003). The same is currently true for ParaHox orthologs in other urochordate genomes: *Ciona savignyi* (CSAV 2.0 at <http://www.ensembl.org>) and *Oikopleura dioica* (V3, <http://www.genoscope.cns.fr>). However, examination of version 2.0 of the *Ciona intestinalis* whole genome shotgun (<http://www.ensembl.org>), in which scaffolds have been mapped to chromosome arms, shows that Gsx is on chromosome 2q, while Xlox and Cdx are found about 250 kb apart on chromosome 14q. The intervening region between Xlox and Cdx contains many predicted or novel genes, transposable elements, and at least one known gene (*Ci-Orphan Fox-1*, UniProt ID Q4H319). Whether this genomic organization can be termed a cluster seems rather doubtful.

In *Ciona intestinalis*, the first ParaHox gene to be expressed is Cdx, in the posterior epidermis and nervous system of the late gastrula (Imai et al., 2004). This is followed by Gsx in the sensory vesicle precursor during the neurula to tailbud stages (Hudson and Lemaire, 2001; Imai et al., 2004), followed by extremely weak expression of Xlox (IPF) in muscle cells in the tailbud stage (Corrado et al., 2001; Imai et al., 2004). Thus the presumed ancestral timings of expression have been rearranged, perhaps having some connection with the dispersal of the cluster (Ferrier and Minguillon, 2003).

In the genomic sequence of the echinoderm *Strongylo-*

\* Corresponding author. Phone: +44-1865-271273;  
Fax : +44-1865-271184;  
E-mail: [rebecca.furlong@zoo.ox.ac.uk](mailto:rebecca.furlong@zoo.ox.ac.uk)

**Table 1.** Gene lengths and intergenic distances in deuterostome ParaHox clusters, expanded and adapted from Ferrier et al. (2005). Lengths are given in base pairs, with the percentage of the total cluster shown in parentheses. The total length of the cluster is measured from the Gsx ATG start site to the Cdx ATG start site. Inferred gene lengths are measured from the ATG start site to the stop codon.

	<i>Homo sapiens</i>	<i>Mus musculus</i>	<i>Monodelphis domestica</i>	<i>Xenopus tropicalis</i>	<i>Amia calva</i>	<i>Eptatretus burgeri</i>	<i>Branchiostoma floridae</i>
<b>Total length of ParaHox cluster</b>	176,315	118,471	197,041	113,674	34,661	57,537	55,826
<b>Gsx gene length</b>	1,257 (0.7%)	1,484 (1.3%)	1,126 (0.6%)	2,021 (1.8%)	1,002 (2.9%)	1,875 (3.3%)	3,280 (5.9%)
<b>Intergenic distance Gsx-Xlox</b>	126,190 (71.6%)	80,244 (67.7%)	145,784 (74.0%)	58,054 (51.1%)	16,075 (46.4%)	–	25,669 (46.0%)
<b>Xlox gene length</b>	4,560 (2.6%)	4,585 (3.9%)	5,469 (2.8%)	24,844 (21.9%)	6,824 (19.7%)	–	8,685 (15.6%)
<b>Intergenic distance Xlox -Cdx</b>	38,416 (21.8%)	27,017 (22.8%)	38,906 (19.7%)	23,122 (20.3%)	7,116 (20.5%)	–	6,947 (12.4%)
<b>Cdx gene length</b>	5,892 (3.3%)	5,141 (4.3%)	5,756 (2.9%)	5,630 (5.0%)	3,644 (10.6%)	19,602 (34.1%)	11,245 (20.1%)

*centrotus purpuratus*, there is a single member of each ParaHox gene family, and again each is dispersed onto a different scaffold with no current evidence for their linkage. Intriguingly, despite this lack of clustering, qPCR reveals that Gsx is expressed first, followed by Xlox (known as *Splox*) and finally Cdx (Arnone et al., 2006). This would be colinear with the presumed ancestral cluster organization as typified by amphioxus, despite being exactly opposite to the amphioxus order of expression. Furthermore, the domains of expression are also colinear with this order of expression and with the amphioxus genomic organization. Very little data are available for other invertebrate deuterostomes. In the hemichordate *Ptychoderma flava*, four ParaHox genes have been sequenced (Gsx, Cdx, and two Xlox [*Lox*] genes) but their genomic organization and expression are unknown (Peterson, 2004). Further analysis of the hemichordate lineage awaits assembly of the *Saccoglossus kowalevskii* genome project currently underway.

In the invertebrate deuterostomes, therefore, it appears that tight clustering of the ParaHox genes may be the exception rather than the norm. While the presumed ancestral order of expression seems to be possible even in the absence of a tight cluster in *S. purpuratus*, in other organisms such as *Ciona intestinalis*, dispersal of the cluster may have occurred in concert with an uncoupling of colinear expression. There are no intervening genes in either the Hox or ParaHox clusters of amphioxus, but transposable elements have been characterised within the ParaHox cluster (Osborne et al., 2006). This suggests that the genomic raw material for rearrangement and relaxation of the cluster is present, but that selective pressure retains the tight cluster in this organism (Ferrier et al., 2005).

#### PARAHOX CLUSTER DUPLICATION

In humans a ParaHox cluster is found on chromosome 13q12.2 (Pollard and Holland, 2000; Ferrier et al., 2005); it contains the genes *GSH1*, *IPF1* (also known as *PDX1*, the ortholog of invertebrate *Xlox*; we use the name Xlox to refer to this gene family in this review since the alternative names imply functions which are not always known) and *CDX2*. While the intergenic distances differ, the order and orientation of the genes is identical to the single cluster of amphioxus and there are no intervening genes.

The human genome also contains additional paralogs of the Gsx and Cdx families (*GSH2*, *CDX1*, *CDX4*), each on a different chromosome. The ParaHox genes of mouse show an equivalent genomic organization (Ferrier et al., 2005). It has long been known that humans and mice possess these

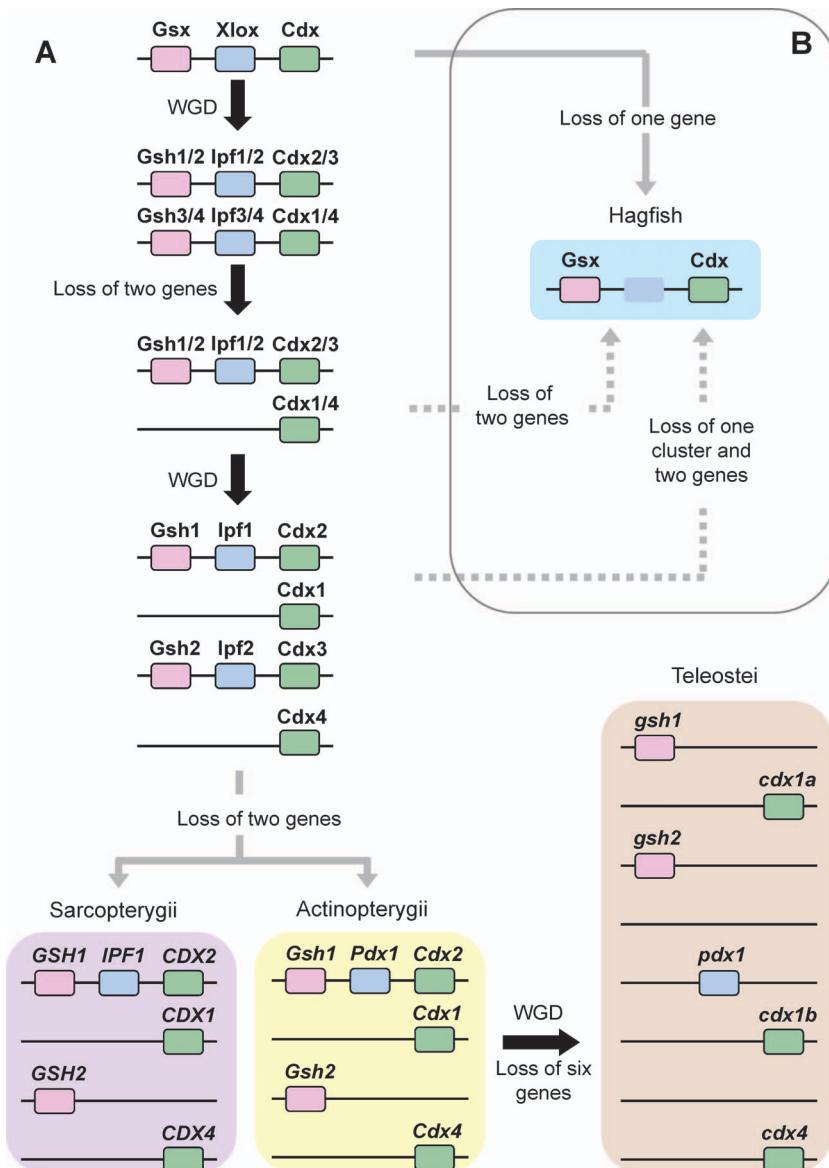
additional members of the Gsx (Singh et al., 1991) and Cdx gene families (Duprey et al., 1988; James and Kazenwadel, 1991; Gamer and Wright, 1993), but not until 2000 was it suggested that they may in fact be the remains of duplicated ParaHox gene clusters Coulier et al., 2000; Pollard and Holland, 2000). Studies of the chromosomal regions containing these genes have revealed that a single prevertebrate ParaHox gene cluster was duplicated en bloc, giving rise to the four ParaHox paralogons seen in the human and mouse genomes (Ferrier et al., 2005). These en-bloc duplications were probably part of two rounds of genome duplication (“2R”) at the base of the vertebrate lineage.

However, neither human nor mouse has four intact ParaHox clusters; rather, they have undergone the loss of at least six genes to retain only a single cluster of *Gsx1*, *Pdx1*, and *Cdx2* with un-clustered *Gsx2*, *Cdx1*, and *Cdx4* spread over three other chromosomes. Our examination of the genome assemblies for the opossum *Monodelphis domestica* (*MonDom5*; Mikkelsen et al., 2007) and frog *Xenopus tropicalis* (v4.1) at <http://www.ensembl.org> show a similar organization, suggesting that these losses occurred prior to the divergence of tetrapods. Mulley et al. (2006) have shown that this organization is also present in basal actinopterygian fish such as *Amia calva* and *Polypterus senegalus*. Therefore we can infer that the common ancestor of all bony fish had undergone both duplications and losses in the ParaHox clusters (Fig. 1).

The complete genome sequences of several species of teleost fish have revealed that the entire teleost lineage is derived from a tetraploid ancestor, the so-called “3R” duplication (Jaillon et al., 2004; Meyer and Van de Peer, 2005). This genome duplication event resulted in the loss of the ParaHox cluster in this lineage (Mulley et al., 2006; Prohaska and Stadler, 2006; Siegel et al., 2007). In all species of teleost examined to date, the usually clustered *cdx2* gene has been lost, and *gsh1* and *pdx1* are on two paralogous chromosomes. Of the remaining ParaHox genes, *gsh2* and *cdx4* are found elsewhere in the genome, while duplicated *cdx1* genes have been retained and may compensate for the loss of *cdx2* in these species (Fig. 1).

#### THE PARAHOX COMPLEMENT OF HAGFISHES

As previously mentioned, it has been hypothesised that two whole-genome duplications (“2R”) occurred on the vertebrate stem after the divergence of cephalochordates and urochordates and before the divergence of bony fish (Dehal and Boore, 2005; Holland et al., 1994; Dehal and Boore, 2005). Furthermore it has been suggested that these



**Fig. 1.** (A) Suggested scheme of ParaHox cluster duplication and loss in the vertebrates, minimizing the number of duplication and loss events. Following 1R duplication of the ancestral cluster, two genes were lost. Following 2R duplication of the remaining genes, two further genes were lost. The 3R duplication in the teleost fish led to cluster breakup and extensive gene loss. WGD, whole-genome duplication. (B) Alternative hypotheses for the origin of the hagfish ParaHox cluster, depending on the timing of hagfish divergence from the vertebrate stem group with respect to the timing of the 2R WGDs.

genome duplications preceded and may have permitted the evolution of the innovations seen in vertebrate developmental patterning (Shimeld and Holland, 2000). The timing of these duplications is therefore key to our understanding of vertebrate development and evolution, and the en-bloc duplication of ParaHox gene clusters may be symptomatic of a larger-scale duplication event.

Only two extant groups diverged around the time of these putative duplications. These are the jawless fish, hagfishes and lampreys. The origins and relationships of these two groups are rather contentious; while molecular phylogenetic analysis strongly suggests that they form a monophyletic sister group to the vertebrates (e.g., Kuraku and Kuratani, 2006; Vienne and Pontarotti, 2006), their morphology argues that they are paraphyletic, and that hagfish form a “craniate” sister group to the true vertebrates, including lampreys (e.g., Forey and Janvier, 1993; Forey, 1984). Furthermore, the timing of their divergences with respect to the 2R duplications is also undetermined. While some

cyclostome gene families appear to have diverged prior to the 2R duplications, other cyclostome genes appear to have undergone 1R or 2R duplications, grouping with vertebrate subfamilies rather than falling basally (e.g., Escriva et al., 2002; Bridgham et al., 2006; Zhang and Cohn, 2006). This suggests that the speciation and duplication events may all have occurred in such quick succession that the phylogenetic signal is impossible to resolve. After genome duplication, the organism must undergo the process of diploidization, during which most gene duplicates are lost and the sequences of the surviving paralogs diverge (Wolfe, 2001). Speciation during this diploidization process may confound phylogenetic analysis (Furlong and Holland, 2002). Genome doubling may also occur by hybridization rather than duplication (allopolyploidy); if one parental species contributes to several different tetraploids, their phylogenetic relationships may again be rather difficult to disentangle.

The lamprey genome project is as yet incomplete, and little is known about the ParaHox genes of these organisms.

However, ParaHox genes from two species of hagfish have recently been characterised (Furlong et al., 2007). In *Eptatretus burgeri*, a *Gsx* and a *Cdx* gene are contained in a single BAC in tail-to-tail configuration. No intervening genes are present between them. In *Myxine glutinosa*, an orthologous *Gsx* gene is linked to a pseudo *Xlox* sequence, suggesting that the functional *Xlox* gene has been lost in both species. Transposable elements are prevalent in the published sequences. The “complete” cluster of *E. burgeri* is around 57 kb, similar in size to that of amphioxus and rather smaller than that of many vertebrates. It has previously been suggested that the ParaHox clusters of amphioxus and mammals maintain not just gene order, but also have similarities in relative gene size and spacing (Ferrier et al., 2005). Expanding previous analyses to include all known deuterostome ParaHox clusters (i.e., those in which linkage has been retained and no intervening genes are found) reveals that the relative intergenic distances do remain fairly constant throughout evolution, with the distance between *Gsx* and *Xlox* being around three times that between *Xlox* and *Cdx* (Table 1; Supplementary Information). Without an *Xlox* gene, these distances obviously cannot be calculated for *E. burgeri*; however, the relative gene lengths can be examined. This shows that relative gene length is rather variable across the chordates, and there are no obvious trends.

While the tight linkage of the hagfish ParaHox cluster shows some similarities with the amphioxus cluster, it also shares some novel features with the vertebrate clusters. For example, a conserved non-coding element associated with the *Gsx* gene in vertebrates (Mulley et al., 2006) is present in hagfishes but not in amphioxus (data not shown). This is in keeping with the suggestion of Woolfe et al. (2005) that such regions are not found in invertebrates.

Loss is prevalent among the ParaHox genes, and in all the vertebrate species studied to date, a maximum of one intact cluster remains. The scheme shown in Fig. 1 indicates that, even if hagfishes diverged after the 2R duplications, they would have a maximum of two intact clusters. A phylogenetic analysis of the hagfish ParaHox genes indicates that in fact the hagfishes did not undergo any genome duplications at the base of the vertebrate lineage (Furlong et al., 2007). However, studies of cyclostome Hox genes suggest that both hagfish and lampreys have multiple Hox clusters (Force et al., 2002; Irvine et al., 2002; Stadler et al., 2004). Furthermore, as mentioned above, molecular phylogenetic analysis of jawless fish is often rather inconclusive. Therefore we must also consider the alternative hypotheses (Fig. 1B), that hagfishes underwent 1R or 2R duplications, followed by more extensive loss.

### PARAHOX CLUSTER MAINTENANCE

The absence of an *Xlox* gene in the *E. burgeri* ParaHox cluster is extremely unusual. Some protostome invertebrate species such as *Drosophila melanogaster* and *Caenorhabditis elegans* lack an *Xlox*, but they also lack any linkage of their ParaHox genes (Ferrier and Holland, 2002). The *Oikopleura* genome also contains no *Xlox* gene (Edvardsen et al., 2005), but again the *Cdx* and *Gsx* genes have not yet been found to be chromosomally linked. It seems that, in deuterostome organisms, the three-gene cluster may be maintained in its entirety, but that any loss of genes is also

associated with dispersal. The close linkage of *E. burgeri* *Gsx* and *Cdx* seems to be the only current example of a tightly linked partial cluster. An obvious question is whether these genes are expressed in a colinear manner; with the recent advances in hagfish embryology (Ota et al., 2007) it may soon become possible to test this.

There are several possible explanations for the maintenance of a ParaHox cluster in animal genomes over hundreds of millions of years. Firstly, the arrangement of the genes in the cluster may affect their spatial and temporal expression during development (as is the case in the Hox gene clusters) if order and location of gene expression were found to follow the order of genes along the chromosome (Ferrier and Minguillon, 2003). Evidence from echinoderms would seem to counter this argument; however, since they do not have a tight cluster but appear to have spatial and temporal colinearity of expression, suggesting that apparent colinearity may be an artefact of the order of tissue development (Frobis and Seaver, 2006; Arnone et al., 2006). Secondly, the ParaHox genes may share overlapping regulatory sequences which trap the genes next to each other, allowing the genes to become dispersed only if these regulatory regions are duplicated as well – this hypothesis is supported by the dispersal of ParaHox genes to different chromosomes in teleost fish following whole genome duplication (Mulley et al., 2006). Finally, the ParaHox genes may be trapped together as “bystander genes” (Becker and Lenhard, 2007; Kikuta et al., 2007) by the action of long range regulatory sequences across them. In this case, there is no functional significance to the presence of *Gsx*, *Xlox*, and *Cdx* next to each other and the cluster is only maintained through the detrimental effects of genomic rearrangement to neighboring genes and their enhancers. This third hypothesis may also explain the break-up of the ParaHox cluster in teleost fish. At the present time there are insufficient data available to be able to choose between these options. More data on the timing and location of ParaHox gene expression in vertebrates are required, as well as a thorough analysis of conserved non-coding sequences associated with the ParaHox cluster which may be acting on it, or across it on adjacent genes. The supplementary information for this article can be found online at <http://dx.doi.org/10.2108/zsj.25.955.s1> .

### ACKNOWLEDGMENTS

We would like to thank Peter Holland for research support, and for helpful discussions and comments on the manuscript. RFF is funded by Christ Church, Oxford and the John Fell Fund.

### REFERENCES

- Arnone MI, Rizzo F, Annunziata R, Cameron RA, Peterson KJ, Martinez P (2006) Genetic organization and embryonic expression of the ParaHox genes in the sea urchin *S. purpuratus*: insights into the relationship between clustering and colinearity. *Dev Biol* 300: 63–73
- Becker TS, Lenhard B (2007) The random versus fragile breakage models of chromosome evolution: a matter of resolution. *Mol Genet Genomics* 278: 487–491
- Bourlat SJ, Juliusdottir T, Lowe CJ, Freeman R, Aronowicz J, et al. (2006) Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* 444: 85–88
- Bridgham JT, Carroll SM, Thornton JW (2006) Evolution of hormone-receptor complexity by molecular exploitation. *Science*

312: 97–101

Brooke NM, Garcia-Fernandez J, Holland PW (1998) The ParaHox gene cluster is an evolutionary sister of the hox gene cluster. *Nature* 392: 920–922

Chourrout D, Delsuc F, Chourrout P, Edvardsen RB, Rentzsch F, et al. (2006) Minimal ProtoHox cluster inferred from bilaterian and cnidarian Hox complements. *Nature* 442: 684–687

Corrado M, Aniello F, Fucci L, Branno M (2001) *Ci-lpf1*, the pancreatic homeodomain transcription factor, is expressed in neural cells of *Ciona intestinalis* larva. *Mech Dev* 102: 271–274

Coulier F, Burtey S, Chaffanet M, Birg F, Birnbaum D (2000) Ancestrally duplicated ParaHox gene clusters in humans. *Int J Oncol* 17: 439–444

Dehal P, Boore JL (2005) Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol* 3: e314

Delsuc F, Brinkmann H, Chourrout D, Philippe H (2006) Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* 439: 965–968

Duprey P, Chowdhury K, Dressler GR, Balling R, Simon D, Guenet JL, Gruss P (1988) A mouse gene homologous to the *Drosophila* gene *caudal* is expressed in epithelial cells from the embryonic intestine. *Genes Dev* 2: 1647–1654

Edvardsen RB, Seo HC, Jensen MF, Mialon A, Mikhaleva J, et al. (2005) Remodelling of the homeobox gene complement in the tunicate *Oikopleura dioica*. *Curr Biol* 15: R12–R13

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

Ferrier DE, Holland PW (2002) *Ciona intestinalis* ParaHox genes: evolution of Hox/ParaHox cluster integrity, developmental mode, and temporal colinearity. *Mol Phylogenet Evol* 24: 412–417

Ferrier DE, Minguillon C (2003) Evolution of the Hox/ParaHox gene clusters. *Int J Dev Biol* 47: 605–611

Ferrier DE, Dewar K, Cook A, Chang JL, Hill-Force A, Amemiya C (2005) The chordate ParaHox cluster. *Curr Biol* 15: R820–R822

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

Forey PL (1984) Yet more reflections on agnathan-gnathostome relationships. *J Vertebr Paleontol* 4: 330–343

Forey PL, Janvier P (1993) Agnathans and the origin of jawed vertebrates. *Nature* 361: 129–134

Frobius AC, Seaver EC (2006) ParaHox gene expression in the polychaete annelid *Capitella* sp. I. *Dev Genes Evol* 216: 81–88

Furlong RF, Holland PW (2002) Were vertebrates octoploid? *Phil Trans R Soc Lond B* 357: 531–544

Furlong RF, Younger R, Kasahara M, Reinhardt R, Thorndyke M, Holland PW (2007) A degenerate ParaHox gene cluster in a degenerate vertebrate. *Mol Biol Evol* 24: 2681–2686

Gamer LW, Wright CV (1993) Murine *Cdx-4* bears striking similarities to the *Drosophila caudal* gene in its homeodomain sequence and early expression pattern. *Mech Dev* 43: 71–81

Garcia-Fernandez J (2005) Hox, ParaHox, ProtoHox: facts and guesses. *Heredity* 94: 145–152

Garcia-Fernandez J, Holland PW (1994) Archetypal organization of the amphioxus Hox gene cluster. *Nature* 370: 563–566

Holland PW (2001) Beyond the Hox: how widespread is homeobox gene clustering? *J Anat* 199: 13–23

Holland PW, Garcia-Fernandez J, Williams NA, Sidow A (1994) Gene duplications and the origins of vertebrate development. *Dev Suppl* 125–133

Hudson C, Lemaire P (2001) Induction of anterior neural fates in the ascidian *Ciona intestinalis*. *Mech Dev* 100: 189–203

Imai KS, Hino K, Yagi K, Satoh N, Satou Y (2004) Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. *Development* 131: 4047–4058

Irvine SQ, Carr JL, Bailey WJ, Kawasaki K, Shimizu N, Amemiya CT, Ruddle FH (2002) Genomic analysis of Hox clusters in the sea lamprey *Petromyzon marinus*. *J Exp Zool* 294: 47–62

Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, et al. (2004) Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431: 946–957

James R, Kazenwadel J (1991) Homeobox gene expression in the intestinal epithelium of adult mice. *J Biol Chem* 266: 3246–3251

Kikuta H, Laplante M, Navratilova P, Komisarczuk AZ, Engstrom PG, et al. (2007) Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Res* 17: 545–555

Kuraku S, Kuratani S (2006) Time scale for cyclostome evolution inferred with a phylogenetic diagnosis of hagfish and lamprey cDNA sequences. *Zool Sci* 23: 1053–1064

Meyer A, Van de Peer Y (2005) From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *BioEssays* 27: 937–945

Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, et al. (2007) Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* 447: 167–177

Mulley JF, Chiu CH, Holland PW (2006) Breakup of a homeobox cluster after genome duplication in teleosts. *Proc Natl Acad Sci USA* 103: 10369–10372

Osborne PW, Luke GN, Holland PW, Ferrier DE (2006) Identification and characterisation of five novel miniature inverted-repeat transposable elements (MITEs) in amphioxus (*Branchiostoma floridae*). *Int J Biol Sci* 2: 54–60

Ota KG, Kuraku S, Kuratani S (2007) Hagfish embryology with reference to the evolution of the neural crest. *Nature* 446: 672–675

Peterson KJ (2004) Isolation of Hox and ParaHox genes in the hemichordate *Ptychodera flava* and the evolution of deuterostome Hox genes. *Mol Phylogenet Evol* 31: 1208–1215

Pollard SL, Holland PW (2000) Evidence for 14 homeobox gene clusters in human genome ancestry. *Curr Biol* 10: 1059–1062

Prohaska SJ, Stadler PF (2006) Evolution of the vertebrate ParaHox clusters. *J Exp Zool B* 306: 481–487

Shimeld SM, Holland PW (2000) Vertebrate innovations. *Proc Natl Acad Sci USA* 97: 4449–4452

Siegel N, Hoegg S, Salzburger W, Braasch I, Meyer A (2007) Comparative genomics of paraHox clusters of teleost fishes: gene cluster breakup and the retention of gene sets following whole genome duplications. *BMC Genomics* 8: 312

Singh G, Kaur S, Stock JL, Jenkins NA, Gilbert DJ, Copeland NG, Potter SS (1991) Identification of 10 murine homeobox genes. *Proc Natl Acad Sci USA* 88: 10706–10710

Stadler PF, Fried C, Prohaska SJ, Bailey WJ, Misof BY, Ruddle FH, Wagner GP (2004) Evidence for independent Hox gene duplications in the hagfish lineage: a PCR-based gene inventory of *Eptatretus stoutii*. *Mol Phylogenet Evol* 32: 686–694

Vienne A, Pontarotti P (2006) Metaphylogeny of 82 gene families sheds a new light on chordate evolution. *Int J Biol Sci* 2: 32–37

Wada S, Tokuoka M, Shoguchi E, Kobayashi K, Di Gregorio A, et al. (2003) A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. II. Genes for homeobox transcription factors. *Dev Genes Evol* 213: 222–234

Wolfe KH (2001) Yesterday's polyploids and the mystery of diploidization. *Nat Rev Genet* 2: 333–341

Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, et al. (2005) Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol* 3: 116–130

Zhang G, Cohn MJ (2006) Hagfish and lancelet fibrillar collagens reveal that Type II collagen-based cartilage evolved in stem vertebrates. *Proc Natl Acad Sci USA* 103: 16829–16833