Available online at www.sciencedirect.com

ScienceDirect

Developmental Biology 300 (2006) 63–73

DEVELOPMENTAL BIOLOGY

www.elsevier.com/locate/ydbio

Genetic organization and embryonic expression of the ParaHox genes in the sea urchin *S. purpuratus*: Insights into the relationship between clustering and colinearity

Maria I. Arnone [a], Francesca Rizzo [a], Rosella Annunciata [a], R. Andrew Cameron [b], Kevin J. Peterson [c], Pedro Martínez [d,e,*]

[a] *Stazione Zoologica 'Anton Dohrn', Villa Comunale, 80121 Napoli, Italy*
[b] *Division of Biology 139-74, California Institute of Technology, Pasadena, CA 91125, USA*
[c] *Department of Biological Sciences, Dartmouth College, Hanover, NH 03755, USA*
[d] *Department de Genètica, Universitat de Barcelona, Av. Diagonal, 645, 08028 Barcelona, Spain*
[e] *ICREA (Institut Català de Recerca i Estudis Avancats), Barcelona, Spain*

Received for publication 30 May 2006; revised 20 July 2006; accepted 27 July 2006
Available online 4 August 2006

Abstract

The ANTP family of homeodomain transcription factors consists of three major groups, the NKL, the extended Hox, and the Hox/ParaHox family. Hox genes and ParaHox genes are often linked in the genome forming two clusters of genes, the Hox cluster and the ParaHox cluster, and are expressed along the major body axis in a nested fashion, following the relative positions of the genes within these clusters, a property called colinearity. While the presences of a Hox cluster and a ParaHox cluster appear to be primitive for bilaterians, few taxa have actually been examined for spatial and temporal colinearity, and, aside from chordates, even fewer still manifest it. Here we show that the ParaHox genes of the sea urchin *Strongylocentrotus purpuratus* show both spatial and temporal colinearity, but with peculiarities. Specifically, two of the three ParaHox genes-discovered through the *S. purpuratus* genome project-*Sp-lox* and *Sp-Cdx*, are expressed in the developing gut with nested domains in a spatially colinear manner. However, transcripts of *Sp-Gsx*, although anterior of *Sp-lox*, are detected in the ectoderm and not in the gut. Strikingly, the expression of the three ParaHox genes would follow temporal colinearity if they were clustered in the same order as in chordates, but each ParaHox gene is actually found on a different genomic scaffold (>300 kb each), which suggests that they are not linked into a single coherent cluster. Therefore, ParaHox genes are dispersed in the genome and are used during embryogenesis in a temporally and spatially coherent manner, whereas the Hox genes, now fully sequenced and annotated, are still linked and are employed as a complex only during the emergence of the adult body plan in the larva.

Keywords: Hox; ParaHox; Colinearity; *Gsx*; *lox*; *Cdx*; Cluster; *S. purpuratus*; Evolution

Introduction

Homeobox genes encode transcription factors that during development regulate the expression of a great variety of downstream genes. In animals, homeobox genes form a large family, with more than 200 members in humans and about 100 in both *Drosophila* and *Caenorhabditis* (Nam and Nei, 2005). Based largely on sequence similarities, these genes have been classified into 49 different families. Among these, the Hox/ParaHox family has, in recent years, received a great deal of attention because of their general importance for both the evolution and development of animal form (Carroll, 1995). Arguably the most important and most interesting feature of this family is that not only are Hox genes and ParaHox genes linked in the genome (forming the Hox and the ParaHox complexes), but that both clusters manifest colinearity. Colinearity is the correlation between the positions of Hox/ParaHox genes with their respective clusters and either their expression domains along the major body axes (spatial) or their transcriptional activation (temporal) such that genes at the 3′ end of the cluster

* Corresponding author. Department de Genètica, Universitat de Barcelona, Av. Diagonal, 645, 08028-Barcelona, Spain. Fax: +34 934034420.
E-mail address: Pedro.martinez@ub.edu (P. Martínez).

0012-1606/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2006.07.037

are expressed first and/or anterior of those located more 5′ (Duboule, 1998). However, recent data from many laboratories suggest that this view might represent an oversimplification because not only are Hox clusters and ParaHox clusters often broken, but few examples of spatial and temporal colinearity are actually described outside of chordates. Here, we explore these propositions with respect to the ParaHox genes in the context of the *Strongylocentrotus purpuratus* genome project.

Hox genes were first identified in sea urchins in 1986 from the Hawaiian species *Tripneustes gratilla* (Dolecki et al., 1986). Since then, Hox genes have been cloned in many different species, mostly using PCR screens. However, only very recently, with the cloning and sequencing of the *S. purpuratus* HOX cluster region, have we been able to determine the exact number (eleven) and the arrangement in the cluster (Cameron et al., 2006; this paper). Although possessing the usual complements of anterior and posterior genes, the Hox cluster of the sea urchin is unusual with respect to other animals in that *Hox4* has been lost, and *Hox1–3* have been translocated to the 5′ end of the cluster in opposite orientation (see Fig. 1); the break point, that between *Hox3* and *Hox5*, may explain the conspicuous absence of the *Hox4* orthologue (Martinez et al., 1999; Long et al., 2003; Cameron et al., 2006). Most sea urchin Hox genes are activated late in development, at the larval stages when the adult body is being generated (Arenas-Mena et al., 1998, 2000; Morris and Byrne, 2005). In fact, only two of the eleven genes are expressed at significant levels during embryogenesis: *Hox7* and *Hox11/13b* (Angerer et al., 1989; Dobias et al., 1996; Arenas-Mena et al., 1998). During the *S. purpuratus* larval stages, when most of the genes in the cluster are being activated, the most posterior genes (*Hox7*, *8*, *9/10*, *11/13a*, and *11/13b*) display a colinear expression pattern within the somatocoels (Arenas-Mena et al., 2000; Fig. 1), reminiscent of the patterns shown by these genes in other bilaterians.

Several Hox genes have an evolutionary sister gene or paralogue, called a ParaHox gene. There are three ParaHox genes: (1) *Gsx*, the paralogue of the anterior Hox genes; (2) *Xlox*, the paralogue of *Hox3*; and (3) *Cdx*, the paralogue of the posterior Hox genes (Brooke et al., 1998). Brooke et al. (1998) made the interesting observation that in amphioxus the ParaHox genes are linked in a manner reminiscent of the Hox genes, with the gene order 3′-*Gsx*-*Xlox*-*Cdx*-5′. Because Hox genes are also linked, this is probably a primitive feature of the ParaHox genes, and suggests that the Proto-Hox genes (i.e., the ancestral genes to both Hox and ParaHox genes) were also linked (Garcia-Fernandez, 2005). However, unlike Hox genes, ParaHox genes seem much more evolutionary labile. For example, both *Drosophila melanogaster* and *Caenorhabditis elegans* lack *Xlox*, and *C. elegans* also lacks *Gsx*, and in the fly *Gsx* (called *Ind*) is not linked with *Caudal*. In the ascidian *Ciona intestinalis*, all three genes are found, but again there is no evidence of clustering (Ferrier and Holland, 2002). Most other taxa examined (e.g., molluscs, annelids, sipunculans, hemichordates) still have all three genes, but no information is available regarding genomic organization (Ferrier and Holland, 2001; Peterson, 2004; Barucca et al., 2006; Frobius and Seaver, 2006).

Brooke et al. (1998) also argued that the ParaHox genes exhibited spatial, but not temporal, colinearity of gene expression. Generally, in those animals in which the spatial expression has been determined, *Gsx* is expressed in or near the anterior regions of the embryo while *Xlox* is mostly expressed in central regions, and *Cdx* expressed at the posterior end (Brooke et al., 1998), although there are exceptions to this rule (Frobius and Seaver, 2006). However, there is discordance in the tissues of expression because *Gsx* is usually expressed in neural tissue while *Xlox* and *Cdx* are expressed in the endoderm. In addition, because few taxa have been examined for clustering, there are actually few examples of true colinearity, although we, and others, often use the expected gene positions, based on orthologous relationships (i.e., the primitive condition), as references when talking about colinearity.

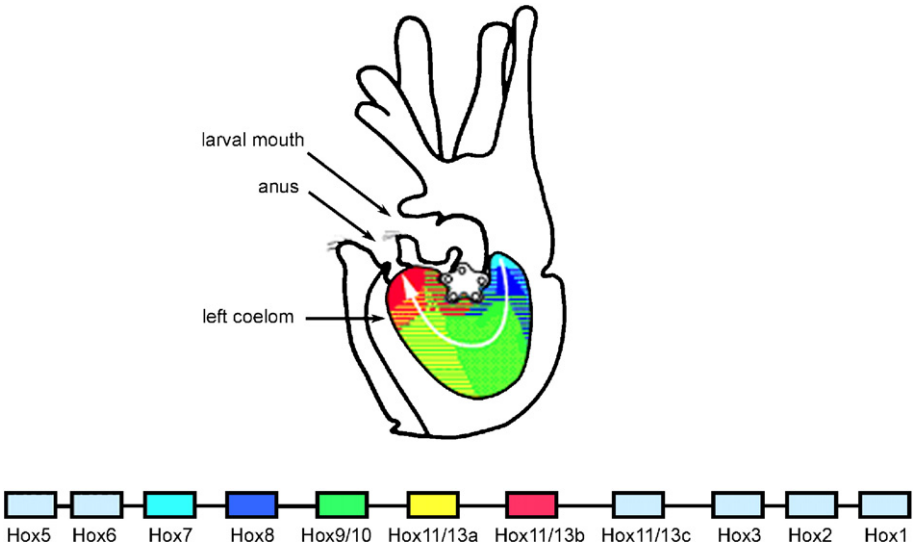

Fig. 1. Diagram of posterior Hox gene expression in the larval somatocoel. The gut assumes a U-shape and the coelom lies against the stomach (obscured in this diagram). The expression domains are depicted using the same colors than the genes, here represented in the sea urchin cluster.

Here, we analyze the genomic position and structure of all three ParaHox genes in the sea urchin *S. purpuratus*, in addition to showing both their transcriptional profiles and spatial localizations during embryogenesis. We show that although these genes are not linked in the genome, they display interesting patterns of spatial and temporal colinearity. Furthermore, when compared with what is already known about Hox gene genomic structure and expression patterns (Arenas-Mena et al., 2000; Cameron et al., 2006), the interesting fact arises that in the sea urchin ParaHox genes are dispersed in the genome and are used during embryogenesis, whereas the Hox genes are still linked and are employed as a complex only during the emergence of the adult body plan in the larva.

Materials and methods

Animals

Adult *S. purpuratus* were obtained from the Kerchoff Marine Laboratory, Caltech, USA. Spawning was induced by intracoelomic injection of 0.5 M KCl and embryos were cultured in seawater in a temperature-controlled incubator (15°C) at the Stazione Zoologica, Naples.

In silico analysis of the Hox and ParaHox genomic region

The genomic sequences spanning all the Hox and ParaHox genes were retrieved from the deposited traces of the *S. purpuratus* genome, at the Baylor site: http://annotation.hgsc.bcm.tmc.edu/. Scaffolds containing all genes were analyzed using the GLEAN3 predictions (a method that includes comparison with known sets of proteins) and GENESCAN algorithm (http://genes.mit.edu/GENSCAN.html) then the predictions were manually annotated. This genomic sequence information was used to predict (and in the known cases, validate) gene features including: intron and exon numbers, intron positions, exon distances, and gene linkages. Similarity searches were done using the BLAST tool and structural motifs were detected using the ScanProSite algorithm.

Figures are provided with all this information compiled. Known cDNA sequences were used as references in the annotation process: *Sp-Hox7* (BAA12813), *Sp-Hox11/13b* (AAB97687), and *Sp-lox* (NP999815).

Quantitative PCR (QPCR)

QPCR was conducted as described by Rast et al. (2000), using a Chromo 4 real-time detector (Biorad, Hercules, CA) and SYBR green chemistry (Applied Biosystems). For the ParaHox genes, a series of specific oligo pairs were used: GSXQF and GSXQR for *Sp-Gsx*, CDXF and CDXR for *Sp-Cdx*, LOXF and LOXR for *Sp-lox*. The oligos were designed based on sequences outside their respective homeoboxes and using the software: Primer3 (http://www-genome.wi.mit.edu/cgi-bin/primer3_www.cgi). The oligonucleotide sequences are:

GSXQF: GAGGATAAGGACGGCATTCA
GSXQR: ACCCTCCTGTTCTGAAACCA
CDXF: AAGGACAAGTATCGCGTCGT
CDXR: CCTTCCGAGAGGCCCAGAG
LOXF: GTGCGACGGACTCCCTATAA
LOXR: TTCAGACGCCATGGTGTAAA

In all QPCR experiments, data from each cDNA sample were normalized against the ubiquitin mRNA and/or 18S ribosomal RNA levels, which are known to remain relatively constant during sea urchin embryogenesis (Nemer et al., 1991). For absolute quantification of the number of transcripts, *SpZ12-1* was used as an internal standard in each cDNA preparation. The number of *SpZ12-1* transcripts in embryos of the relevant stages had been measured earlier by RNA titration (Wang et al., 1995).

Because there are some differences in the absolute transcript levels between different batches of embryos, measurements were always done in duplicates on, at least, three different batches.

Fixation, whole-mount in situ hybridization, and imaging

The whole-mount *in situ* hybridization protocol used here is based on the method described by Minokawa et al. (2004). The accuracy of the whole-mount in situ hybridization results was confirmed in the control experiments using sense probes (not shown). The antisense and sense probes were transcribed in the presence of digoxygenin-11-UTP from appropriate plasmids using a Roche (Indianapolis, Indiana) kit. Images shown in Fig. 5 were made with a Zeiss digital camera (Axiocam) mounted on a Zeiss Axioimage 2 MOT microscope operating in DIC mode. The three probes used were: for *Sp-lox*, a fragment of 800 bp derived from the cDNA clone p16I16ES; for *Sp-Gsx*, a probe was derived from a PCR fragment, cloned in pCRII-TOPO (Invitrogen, Carlsbad, CA), covering a total of 1276 bp of the GLEAN3 predicted sequence (including a 972-bp-long intron), using the oligos GSX QF and GSX DW; for the *Sp-Cdx* gene, we also used a similarly cloned PCR-derived probe containing 1414 bp of genomic sequence encompassing a 829-bp-long intron (as predicted by the GLEAN3 algorithm), using the oligos CDXF and CDXR2. The sequences of CDXF and GSXDW oligos are:

CDXF: AAGGACAAGTATCGCGTCGT
GSXDW: CATTGTCGTCGAAGTCTCCA

The sequences of all PCR fragments were coincident with the predictions obtained from the *S. purpuratus* genome sequence.

Phylogenetic analysis

Sequences were edited and aligned with MacVector 7.0. Sixty amino acids (aa) from the homeodomain of 98 different ANTP genes from *Strongylocentrotus*, a protostome (usually the fly, but when the gene was lost [e.g., *Hox3*] or highly modified [e.g., *tinman*], potential orthologues were taken from other taxa), and the cnidarian *Nematostella vectensis*. All ANTP genes were found in each of the bilaterian taxa, but two appear to be missing from *N. vectensis* (and indeed from all cnidarians), *engrailed*, and *Tlx* (see also Kamm and Schierwater, in press). All sequences were downloaded from Genbank, and only a single member of each potential orthtology group was chosen for analysis (i.e., taxon-specific paralogues were not analyzed). These 98 ANTP genes were analyzed with PAUP v. 4.0b10 for Macintosh (Swofford, 2000) using *Lim* from each taxon to root the final topology. Distance analysis used minimum evolution as the optimality criterion (heuristic search with tree-bisection-reconnection and random addition sequence with 100 replications), and mean character difference as the distance measure. Bootstrap analysis used 1000 replicates.

Results

Genomic structure of the ParaHox genes in S. purpuratus

We found that the genome of *S. purpuratus* contains homologues of the three known ParaHox genes, *Gsx*, *Xlox*, and *Cdx* (Brooke et al., 1998). This is the first time that the presence of ParaHox genes has been reported in sea urchins. In Fig. 2, we describe their homeodomains and those of their closest bilaterian relatives, and Fig. 3 shows a phylogenetic reconstruction of all ANTP genes found in the sea urchin genome as well as their protostome and cnidarian orthologues. *Sp-Gsx*, *Sp-lox*, and *Sp-Cdx* are clearly the orthologues of *Gsx*, *Xlox*, and *Cad* as found in other taxa (Fig. 3 and data not shown). The amino acid sequences for the three genes are provided in the Supporting online material.

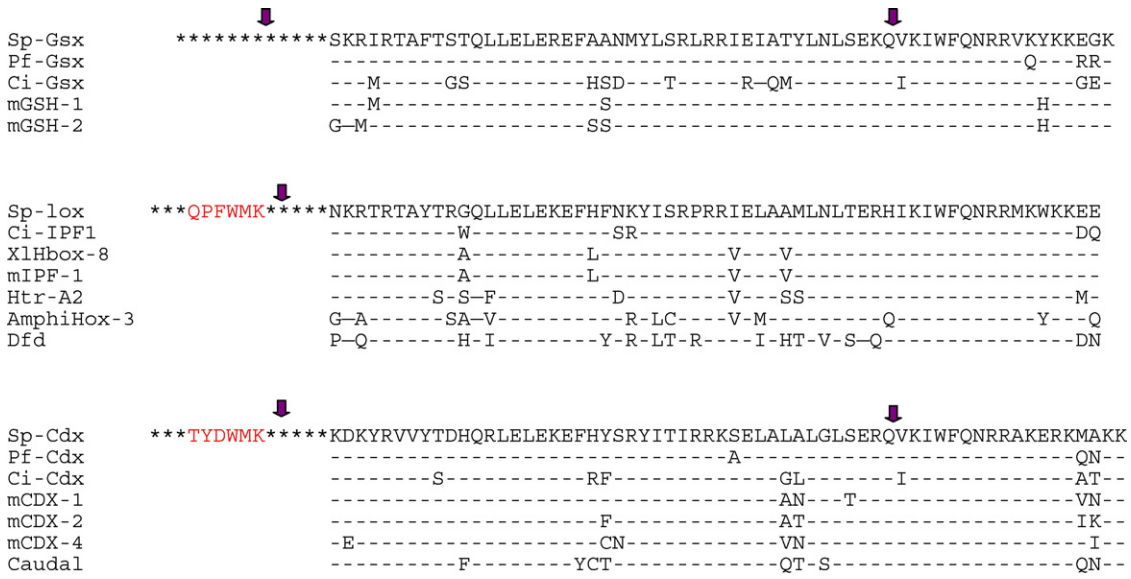

Fig. 2. Alignment of homeodomain sequences for the three *S. purpuratus* ParaHox genes and representative vertebrate and invertebrate orthologues. Broken lines indicate amino acid identities. Dotted lines indicate where the linkers of variable length would be. The hexapeptides are shown in red text. The positions of introns are indicated with a violet vertical arrow. The species names are: Ci (*Ciona intestinalis*); Pf (*Ptychodera flava*); m (mouse); Xl (*Xenopus laevis*); Ht (*Helobdella triserialis*); Amphi (amphioxus, *Branchiostoma floridae*). Dfd and Caudal are *Drosophila melanogaster* genes.

While the ParaHox genes are linked in a genomic cluster in cephalochordates and vertebrates, the sequencing of *S. purpuratus* genome has not revealed a similar organization in sea urchins. Using data provided by the 2006-06-15 assembly of the *S. purpuratus* genome, we have been able to map the positions of the three ParaHox genes within three different large scaffolds (>300 kb, each). Even though the region where they are located comprises now more than 900 kb of genomic sequences, there is no evidence of any linkage among these three genes. Moreover, none of the fragments containing the ParaHox genes also contains the genes *PRHOXNB* or *CHIC*, which are part of a syntenic group well conserved in chordates (Ferrier et al., 2005). The mechanisms that underlie this breakage, and even where it occurred in the evolutionary lineage leading to *S. purpuratus*, are unknown. Because these genes are no longer linked in the genome of *S. purpuratus*, we describe these genes as constituting the ParaHox *group* and not the ParaHox cluster.

Both *Sp-Gsx* and *Sp-Cdx* genes, but not *Sp-lox*, contain introns within the homeobox (indicated by arrows in Fig. 2), and all three genes have a single intron 5′ of the homeobox (Fig. 2); this intron lies between the homeobox and the hexapeptide in *Gsx* and *Cdx*. This exon–intron structure is similar to what is known for other bilaterian ParaHox genes (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=homologene). Interestingly, the position of intron within the homeobox for both *Gsx* and *Cdx* is similar to the position of the intron in their Hox paralogues, *Hox1* and *Hox9/10*, respectively (Fig. 4, and see below). Data regarding gene size, number of exons, and the number of encoded amino acids, are summarized in Table 1.

Some new features of Hox gene structures

Although the Hox gene cluster in the *S. purpuratus* genome is known from completely finished sequence (Martinez et al., 1999; Cameron et al., 2006), we still lacked many details of the genomic structure of these genes. Some key characteristics that would allow us to compare these genes with other bilaterian relatives, such as paralogue group (PG) specific residues, intron positions, and motif conservation outside the homeodomain, have never been fully described. Here, and with the aid of the genomic sequence data and new ESTs obtained through the sea urchin genome project, we are able to investigate those characteristics for all these genes. Our main source of information has been the collection of exon/intron predictions generated through the use of the GLEAN3 and GENESCAN (Burge and Karlin, 1997) algorithms. In the few cases for which we had cDNA sequences available (Angerer et al., 1989; Dobias et al., 1996), these have been used to validate the predictions. Other sea urchin or echinoderm sequences have been brought into our analysis for the further testing of computational predictions. For all Hox genes, GLEAN3 and GENESCAN predictions were identical. Some general difficulties are known to be associated with the prediction of 5′ and 3′ untranslated regions. For that particular reason, we have avoided these in our analysis. This means that gene sizes, for instance, represent minimal approximations to the real (mRNA) size, and describe only those portions covered by the Open Reading Frames. We know that at least for some Hox genes, non-coding sequences are quite long (for instance, *Sp-Hox7*: 5 kb, Martinez et al., 1997), which means that the total size (from the CAP site to the polyadenylation signal) could be much larger than the finally translated portion. The predicted gene size, number of exons, and number of encoded amino acids for each Hox gene are also summarized in Table 1; we emphasize that when cDNA or EST sequences were known, the predictions were all correct.

Sequencing of the full ORFs for the different Hox genes has allowed us, for the first time, to determine the whole amino acid sequence of these proteins and, in particular, the complete

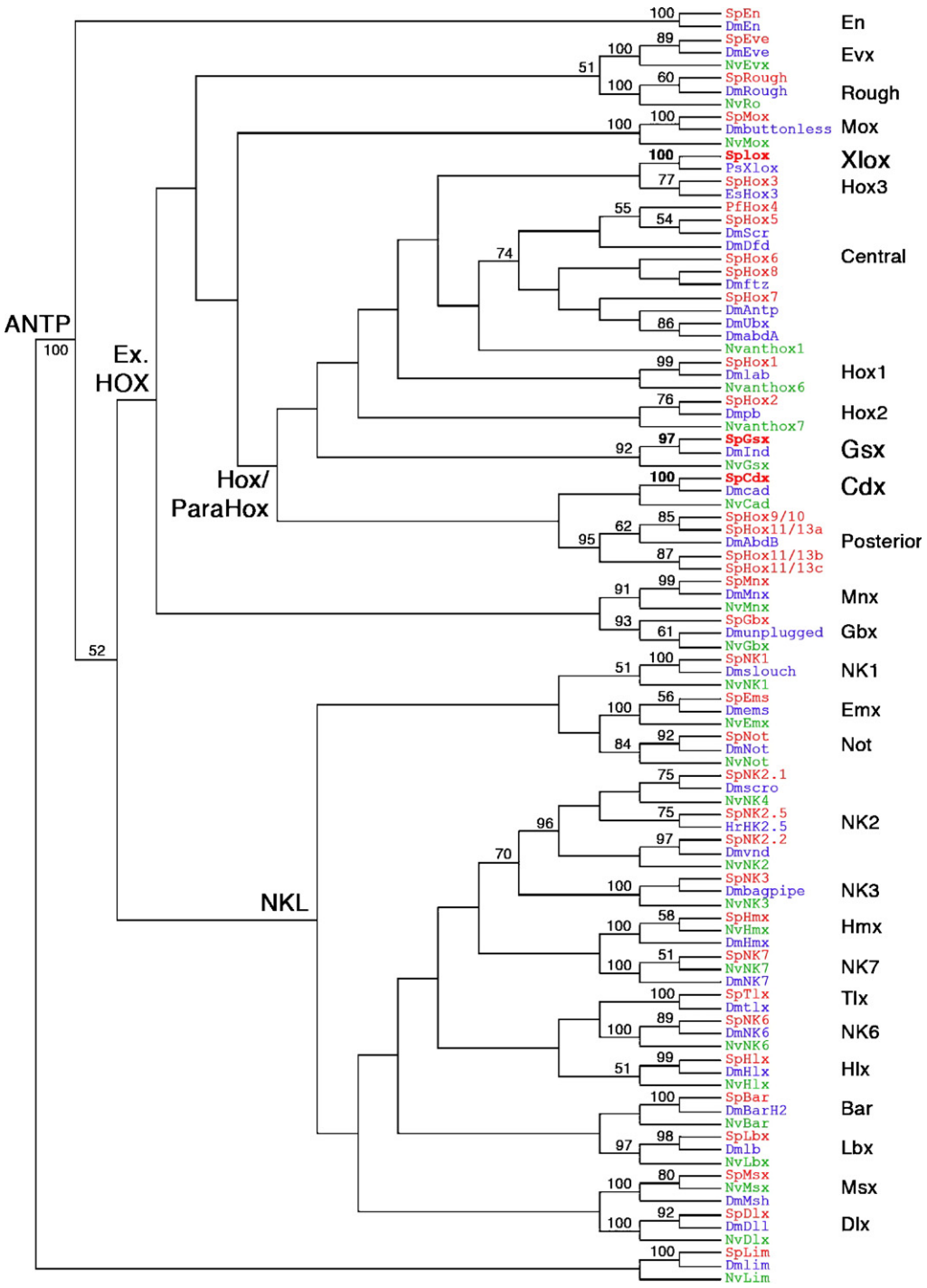

Fig. 3. Phylogenetic analysis of all ANTP genes from the sea urchin *S. purpuratus* red), a protostome (blue), and the cnidarian *Nematostella vectensis* (green) rooted on *Lim*. The ParaHox genes from *S. purpuratus* are shown in bold. Bootstrap values >50% are shown at the respective nodes (derived from 1000 replicates)—note that ParaHox designations of the sea urchin genes are highly supported by bootstrap analysis (indicated in bold). Taxonomic abbreviations are as follows: Dm—*Drosophila melanogaster* (fruit fly); Es—*Euprymna scolopes* (squid); Hr—*Haliotis rufescens* (red abalone); Nv—*Nematostella vectensis* (anthozoan cnidarian); Ps—*Phascolion strombus* (sipunculid worm); Sp—*Strongylocentrotus purpuratus* (sea urchin).

homeodomains. The patterns of paralogue-specific residues seem to be well conserved in all *S. purpuratus* Hox genes (Sharkey et al., 1997; Fig. 4). Indeed, these genes encode for typical Hox proteins, with well-conserved homeodomains typical for their specific orthology groups, and retain conserved motifs, including the hexapeptide domain and other less conserved domains such as the N-terminal peptides (not shown). Alignment of these sequences may be of great value for both phylogenetic purposes and as indicators of regions that have old conserved functions or recently acquired ones.

Intron positions have also been determined for each of the Hox genes in *S. purpuratus* (Table 1, Fig. 4). Most genes have at

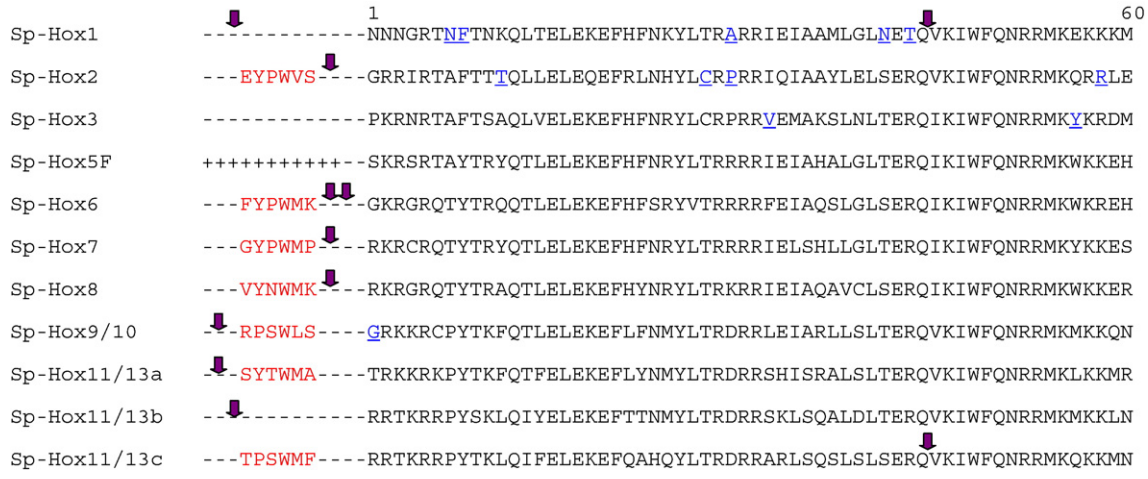

Fig. 4. Alignment of all *S. purpuratus* homeodomain sequences in groups PG1-11/13c. Broken lines indicate where the linkers of variable length would be. The hexapeptides are shown in red text. Paralog-specific residues within the homeodomain are highlighted in blue. The positions of introns are indicated with a violet vertical arrow. Note that *Sp-Hox5* appears fused to a different protein (a consequence of the HOX cluster rearrangement). The fused peptide is marked with cross signs.

least two exons, except for *Hox3*, which appears to lack any introns. Curiously, *Hox5* also lacks introns, but this might be an indirect consequence of the reorganization that this cluster has experienced (Cameron et al., 2006) since the sequence 5′ of the homeobox appears to be from another gene (indicated with crosses on Fig. 4), and thus a gene fusion occurred with the translocation of *Hox1-3* and the loss of *Hox4* (see Fig. 1). The number (Table 1) and position (Fig. 4) of introns are in general agreement with what is known for other bilaterian Hox genes (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=homologene). As mentioned above, *Sp-Hox1* and *Sp-Hox11/13c* contain introns within the homeobox, in agreement with that found in other bilaterian PG1 and Posterior Hox genes (Bürglin, 1995), as well as their respective ParaHox genes (Fig. 2). However, the other posterior genes in the *S. purpuratus* genome (*Sp-Hox9/10*, *Sp-Hox11/13a* and *Sp-Hox11/13b*) seem to have lost this specific intron. Some *Hox* genes (*Sp-Hox2*, *Sp-Hox6*, *Sp-Hox7* and *Sp-Hox8*) possess an intron between the hexapeptide and the homeodomain, but not all genes with a hexapeptide possess this intron (*Sp-Hox9/10*, *Sp-Hox11/13a*, and *SpHox11/13c*) (see Fig. 4).

Table 1
Compilation of gene characteristics for all Hox and ParaHox genes in the genome of *S. purpuratus* [a]

Gene	Size (bp)	Exon (predicted)	CDS (aa)
Hox			
Sp-Hox1	1759	3	182
Sp-Hox2	1821	2	300
Sp-Hox3	728	1	242
Sp-Hox6	19,499	3	321
Sp-Hox7	6291	2	308
Sp-Hox8	13,227	2	305
Sp-Hox9/10	2277	2	92
Sp-Hox11/13a	13,953	2	345
Sp-Hox11/13b	17,427	2	339
Sp-Hox11/13c	145,055	2	150
Para Hox			
Sp-Gsx	9627	3	304
Sp-lox	26,999	2	390
Sp-Cdx	12,021	3	399

[a] Data are based on predictions made with the GLEAN3 and GENESCAN algorithms. The size reflects the span of genomic DNA covered by the ORFs; 3′ and 5′ UTRs are not included given the difficulties of their computational prediction.

Temporal expression of ParaHox genes during embryogenesis

In order to understand how these genes are used during sea urchin embryogenesis, we have analyzed the temporal expression profile for the three genes using quantitative PCR (Fig. 5) (see Materials and methods). Using the expression of the gene *SpZ12-1* as an internal standard (Wang et al., 1995), we have determined the number of transcripts per embryo for each gene at different developmental times. All values have been normalized using the ubiquitin gene (which is known to remain at constant levels over the period analyzed; Nemer et al., 1991) as a reference. The experiments were repeated three times and each time derived from, at least, three different embryo pools.

None of the ParaHox genes encode maternal transcripts, and the activation of each of the three genes peaks at different times during gastrulation. *Sp-Gsx* is detectable from 24 h, reaching a maximum accumulation of transcript at 48 h. After this, levels decrease progressively and reach minimum levels at 64 h. From this time on, the levels increase again and peak at 80 h, this time with about half the number of transcripts present at 48 h. *Sp-lox* is activated later (32–40 h post fertilization) and the mRNA levels accumulate steadily, reaching maximum levels at around 72 h. The levels remain stable for the next 10 h and then drop abruptly during the following eight hours. *Sp-Cdx* levels seem to accumulate steadily from the point of its activation, around 40 h post fertilization, until 80 h. After this, as with *Sp-lox*, levels of *Sp-Cdx* start to steadily decrease until the last time point analyzed at 96 h.

In order to understand the relative pattern of temporal activation for these three genes more clearly, we have also plotted the slope changes in the accumulation curves of each gene (see

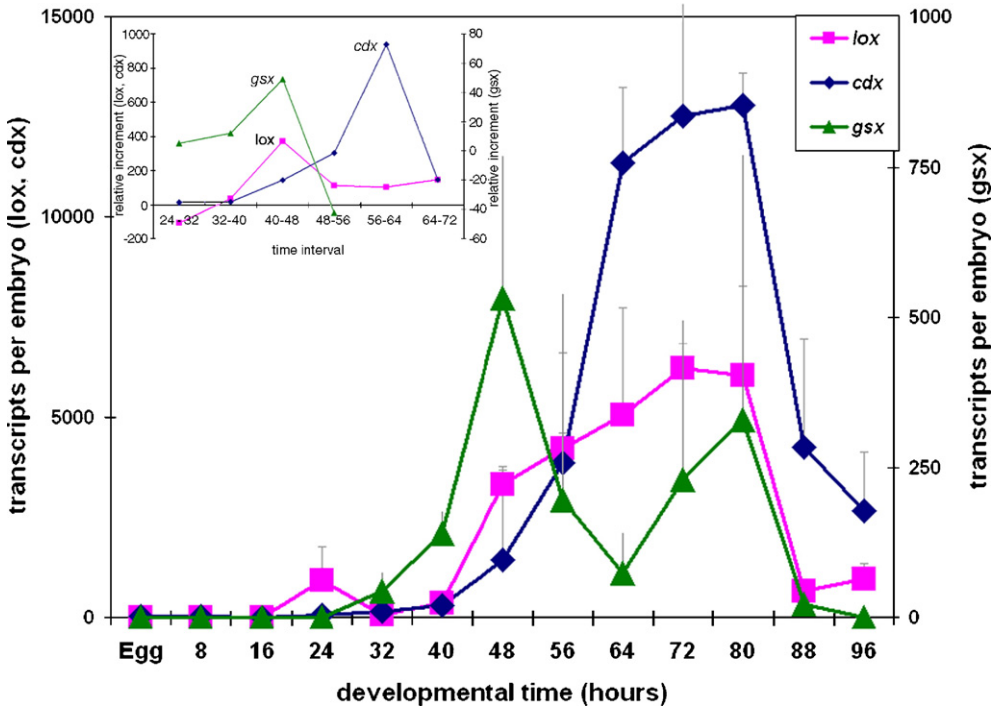

Fig. 5. Temporal expression of the three ParaHox genes during sea urchin development. The graph shows the number of transcripts of *Sp-Gsx*, *Sp-lox*, and *Sp-Cdx* obtained by Quantitative PCR analysis. Conversion of QPCR data to absolute number of transcripts was made using the known amounts of *SpZ12-1* at 24 h as standard. The QPCR values were obtained by comparison of the cycle number in each given reaction required to achieve a threshold set in the exponential phase of the amplification process, and were initially normalized to the ubiquitin mRNA levels present in each sample. The efficiency of amplification was assumed to be 1.9. Given that there are some differences in the absolute transcript levels between different batches of embryos, measurements were always done in triplicate on, at least, three different batches. We have represented also (see inset) the slope changes corresponding to all the three ParaHox genes accumulation profiles.

inset diagram in Fig. 5). This analysis of variation shows a clear successive activation of the three ParaHox genes, following the order: *Sp-Gsx*, then *Sp-lox*, and finally *Sp-Cdx*. This would conform to the so-called temporal colinearity pattern, as defined for Hox genes, but we stress that the term is somewhat inappropriate here given that the genes are no longer clustered.

The spatial domain of expression of ParaHox genes

The spatial expression domain of all ParaHox genes has been determined using whole mount in situ hybridization techniques, following the method described in Minokawa et al. (2004). Gene-specific sequences have been used as hybridization probes, all designed to avoid cross hybridization with putatively related genes. The results are summarized in Fig. 6. In all cases, late gastrulae and plutei embryos are shown because it is at these stages when the expression levels of the ParaHox genes reach their maximum (at different times).

Sp-Gsx is clearly detected at gastrula through pluteus stages, in agreement with the Q-PCR data reported above (Figs. 6A, B). Its expression domain is confined to two small patches of ectodermal cells, one in each side of the embryo, apparently located in the vegetal half of the embryo at gastrula stage and more clearly at the level of the midgut in plutei. Each patch may contain only one to three cells, and the two spots are always of different sizes-consistently, the spot on the right side is always larger than the one on the left. In addition, each of these cells has a small protrusion facing the blastocoelar space (not shown). These three facts are consistent with these cells being neurons, although there are no known neural patterns that conform to this architecture. Nonetheless, there is some resemblance between these cells and those making the ganglion of the lower lip in *Hemicentrotus pulcherrimus* (Yaguchi and Katow, 2003), and the asymmetry in the number of cells could be due to the presence of asynchronously dividing neuroblasts (Bisgrove and Burke, 1986). In any case, double labeling with known markers will be necessary to assess more precisely the cellular affinity of these Gsx-positive cells.

In contrast to the ectodermal expression of *Sp-Gsx*, both *Sp-lox* (Figs. 6C, D) and *Sp-Cdx* (Figs. 6E, F) are expressed during embryogenesis in the growing archenteron (gut). From 72–96 h (late pluteus stage), *Sp-lox* is expressed around the midgut/hindgut sphincter region, whereas *Sp-Cdx* marks a region posterior to the *Sp-lox* domain, in the hindgut and close to the blastoporal opening. The *Sp-Cdx* domain seems to fade anteriorly and overlaps with the *Sp-lox* domain. These patterns resemble the endodermal expression of *Xlox* and *Cdx* genes seen in other bilaterians including the overlap of *Xlox* and *Cdx* expression in the development of the mouse gut (Fang et al., 2006). However, it has been noted that the expression of *Xlox* is not always strictly confined to the gut, although the posterior endodermal domain of *Cdx* is highly conserved. Nonetheless, the consistent expression of *Xlox* and *Cdx* in the bilaterian gut suggests that they might be key components of the regulatory network "kernel" mediating the regionalization of bilaterian guts (Walters et al., 1997; Davidson and Erwin, 2006).

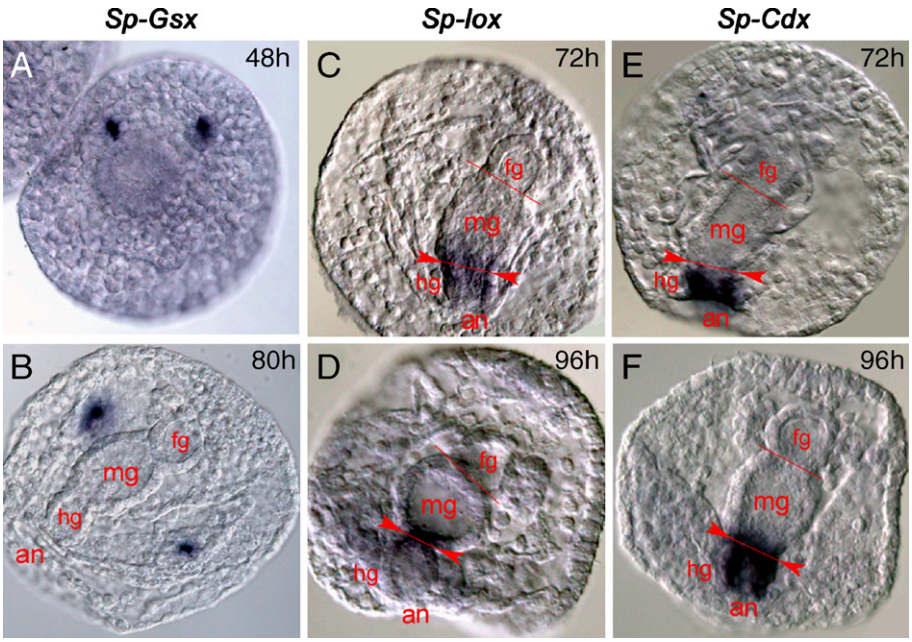

Fig. 6. Spatial expression domains of all three *S. purpuratus* ParaHox genes as determined by whole mount *in situ* hybridization. All staining correspond to gastrulating and larval stages. Panels A and B correspond to the expression of *Sp-Gsx* (48 and 80 h, respectively); panels C and D to the *Sp-lox* expression (72 and 96 h), and panels E and F provide the expression domains for *Sp-Cdx* (72 and 96 h). Arrowheads delimit the midgut–hindgut boundary. Abbreviations: an (anus); fg (foregut), mg (midgut), and hg (hindgut). Panel A is a vegetal view of a gastrula embryo, while panels B, C, D, E, and F are all showing lateral views of plutei of different age. The subdivisions of the tripartite gut are marked with thin, red lines.

These data are consistent with other studies that show a clear distinction between the usage of *Gsx* (neurectoderm) and *Xlox* plus *Cdx* (mainly endoderm) during embryogenesis. In addition, our data provide further support for the notion that some ParaHox genes show spatial colinearity, at least within the gut.

The extended HOX cluster (The MEGACLUSTER)

All Hox and ParaHox genes are members of a larger family of homeobox genes, the so-called ANTP family. This family includes the Hox, ParaHox, Meox, and Gbx orthologues (extended Hox), as well as the NK-like genes and engrailed. All members of this family encode very similar homeobox sequences, and are clearly monophyletic, indicating a common ancestry (Fig. 3). Because the family is monophyletic and retains signatures of clustering, it appears that all ANTP genes were originally members of a single cluster that has been named: "Extended HOX cluster" or MEGACLUSTER (Coulier et al., 2000; Pollard and Holland, 2000; reviewed in Garcia-Fernandez, 2005).

In order to analyze the composition and structure of the Extended Hox cluster (or what remains of it) in sea urchins, we have searched the *S. purpuratus* genome for all ANTP family members. Interestingly, none seem to be missing from the sea urchin genome (Fig. 3). Moreover, there is still evidence for some genomic associations between ANTP-group genes: one is the obvious case of the HOX cluster together with another more surprising linkage between *Meox* and *Gbx*. These two genes appear in a genomic region of around 300 kb. Strikingly, based on current models and the phylogenetic topology (Fig. 3), this is an unexpected association.

The linkage of ANTP homeobox genes has not been reported in other deuterostomes outside vertebrates and amphioxus apart from the larvacean *Oikopleura dioica*, where there is clear evidence of linkage between *Cdx* and *Hox1* (Seo et al., 2004). The presence of other associations, such as the NK cluster, found in *Drosophila* and humans (Kim and Nirenberg, 1989; Pollard and Holland, 2000) supports the evolutionary existence of such a Megacluster of homeoboxes in a bilaterian ancestor. Here we have added further evidence of clustering between members of the ANTP family.

Discussion: does "colinearity" always rule?

The completed characterization of the ParaHox genes in the sea urchin *S. purpuratus* now allows us to compare their structure, function, and evolution, not only with the Hox genes of the sea urchin, but also with the Hox and ParaHox genes amongst other bilaterians.

It is often assumed, based on a small number of taxa, that Hox and ParaHox genes are clustered in the genomes of all bilaterian groups. While this statement is sometimes true for Hox genes, and surely is the primitive condition for both Hox and ParaHox genes, if not the entire ANTP family (see above), there are few examples of ParaHox clustering in bilaterians. Clustering is very important mechanistically because, as was clear from the very first studies on Hox gene regulation, gene order within the cluster and their spatial and/or temporal expression domains were related (Kauffman et al., 1978; Lewis,

1978). Later studies clearly revealed precise colinear relationships between gene positions in the Hox cluster and the expression patterns in the different embryonic layers (reviewed in McGinnis and Krumlauf, 1992; Martinez and Amemiya, 2002). Maybe more interesting though is the increasing number of examples where spatial colinearity is still seen despite the fact that the genes are no longer clustered on the chromosome (Aboobaker and Blaxter, 2003; Negre et al., 2003; Seo et al., 2004). This has led to some terminological confusion with the use of word colinearity, where instead of referring to the relationship between domains of expression and position within a cluster, the term is used also to denote correspondences between these expression domains and the primitive cluster position, based on orthologous relationships of the genes under consideration, independently of the actual knowledge of the cluster structure. Nonetheless, elucidating the mechanisms underlying this spatial "colinearity" when the genes are no longer linked remains an exciting challenge to more fully understand Hox and ParaHox gene function.

In addition to spatial colinearity, a functionally important relationship was revealed in the progressive temporal activation of Hox genes during mammalian embryonic development and in cultured mammalian cells (Dolle et al., 1989; Simeone et al., 1990). Both aspects of colinearity might be mechanistically linked (at least in some groups), with the timing of gene initiation determining the axial limits of expression (Duboule, 1994). In fact, it is possible that the constraints imposed by this need of sequential (temporal) expression in vertebrates have kept the HOX genes tightly bound. However, temporal colinearity is not (generally) conserved in invertebrates. Neither Drosophilids nor *C. elegans* show temporal colinearity (McGinnis and Krumlauf, 1992; Aboobaker and Blaxter, 2003), but neither possesses a complete and intact Hox cluster either. Thus, although temporal colinearity seems to require the presence of a tight HOX arrangement, spatial colinearity does not, at least among some invertebrates.

Brooke et al. (1998) proposed that the ParaHox group may also follow colinearity rules similar to those exhibited by Hox genes, and this appears to be the case at least in chordates (Ferrier and Holland, 2002). Although in the sea urchin the orthologues of *Gsx*, *Xlox*, and *Cdx* are clearly present, there is no evidence for any physical linkage among them. Nonetheless, some colinearity rules might still be followed. For instance, with respect to temporal colinearity, the activation of the three genes seems to follow the order of the genes within the ancestral ParaHox cluster, starting with *Sp-Gsx* and finishing with *Sp-Cdx*, as is clearly demonstrated by our Q-PCR experiments (Fig. 5). Obviously, because the genes show no apparent linkage, the mechanism(s) that produces such a regulation cannot operate in *cis*. Some authors have suggested that the timing of expression simply reflects the time at which the different tissues (gut and neural) expressing ParaHox are specified or differentiated (Frobius and Seaver, 2006), and thus this temporal correlation between chromosomal order and transcription activation is purely coincidental. Whether this hypothesis can be generalized to other taxa remains unknown given that it was erected on the basis of ParaHox gene usage in a taxon with a non-feeding larval form.

Strikingly, analysis of the spatial pattern of expression for the ParaHox genes shows that two (*Sp-lox* and *Sp-Cdx*) are expressed in staggered domains within the gut primordium with *Sp-lox* anterior of *Sp-Cdx*. The third gene, *Sp-Gsx*, is expressed more anteriorly, but in the ectoderm. A mechanistic question arises regarding how these nested spatial domains are initiated and maintained. So far, all we can propose is that cross-regulatory interactions between *Xlox* and *Cdx* are likely involved. In this context, it would be of capital importance to understand whether ParaHox genes are actually used to pattern the different gut regions, and how their activities are affected by each other. For example, proper regionalization of the gut absolutely depends upon the activity of the *Xlox* gene, as a MASO directed against *Sp-lox* dramatically affects proper gut development (M.I. Arnone and P. Martinez, unpublished observations). Because of the ability to knock-down the translation of these genes in the context of relatively simple embryo with a very well characterized GRN for gut development (Davidson et al., 2002), the sea urchin embryo offers us an unprecedented opportunity to look at cross-regulation among ParaHox genes, an issue never approached in any other biological system.

Here we have considered two different sets of genes, Hox and ParaHox, which operate during sea urchin development. These genes encode proteins with conserved homeodomains typical for their orthology groups, and retain conserved motifs, including the hexapeptide domain, among family members. In both cases, the genes show some aspects of spatial colinearity. Temporal colinearity is not respected by the Hox genes, however, it is respected by the ParaHox genes. Perhaps the most interesting and important difference between the Hox and ParaHox genes of the sea urchins concerns their usage during development. The Hox genes are used primarily during the larval stage (with some apparent cooption during embryogenesis) in cells destined to become parts of the juvenile body plan, whereas the ParaHox genes function during embryogenesis in the context of gut and ectoderm development. It will be most interesting to observe how these genes are used in other indirect developing invertebrates, especially other deuterostomes including hemichordates.

Acknowledgments

We thank Michael Thorndyke (Kristineberg Marine Station, Sweden) and Jordi Garcia-Fernández (Universitat de Barcelona, Spain) for their very helpful comments on the manuscript. We would also like to thank Marco Borra and Elvira Mauriello of the SZN Molecular Biology Service for their assistance with QPCR and sequencing. We are very grateful to the European Network of Excellence "Marine Genomics Europe" (GOCE-04-505403) for supporting the organization of the *S. purpuratus* Annotation Workshop, held at the Stazione Zoologica di Napoli, Italy, in January 2006. National Science Foundation Grant IBN9604454 and the Beckman Institute, California Institute of Technology supported the work on this project contributed by R. A. Cameron. NASA-Ames, NSF, and the Dartmouth College support K. J. Peterson.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ydbio.2006.07.037.

References

Aboobaker, A.A., Blaxter, M.L., 2003. Hox gene loss during dynamic evolution of the nematode cluster. Curr. Biol. 13, 37–40.

Angerer, L.M., Dolecki, G.J., Gagnon, M.L., Lum, R., Wang, G., Yang, Q., Humphreys, T., Angerer, R.C., 1989. Progressively restricted expression of a homeo box gene within the aboral ectoderm of developing sea urchin embryos. Genes Dev. 3, 370–383.

Arenas-Mena, C., Martinez, P., Cameron, R.A., Davidson, E.H., 1998. Expression of the Hox gene complex in the indirect development of a sea urchin. Proc. Natl. Acad. Sci. U. S. A. 95, 13062–13067.

Arenas-Mena, C., Cameron, A.R., Davidson, E.H., 2000. Spatial expression of Hox cluster genes in the ontogeny of a sea urchin. Development 127, 4631–4643.

Barucca, M., Biscotti, M.A., Olmo, E., Canapa, A., 2006. All the three ParaHox genes are present in *Nuttallochiton mirandus* (Mollusca: polyplacophora): evolutionary considerations. J. Exp. Zool. B Mol. Dev. Evol. 306, 164–167.

Bisgrove, B.W., Burke, R.D., 1986. Development of serotonergic neurons in embryos of the sea urchin, *Strongylocentrotus purpuratus*. Dev. Growth Differ. 28, 569–574.

Brooke, N.M., Garcia-Fernandez, J., Holland, P.W., 1998. The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. Nature 392, 920–922.

Burge, C., Karlin, S., 1997. Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. 268, 78–94.

Bürglin, T.R., 1995. The evolution of homeobox genes. In: Arai, R., Kato, M., Doi, Y. (Eds.), Biodiversity and Evolution. The National Science Museum Foundation, Tokyo, pp. 291–336.

Cameron, R.A., Rowen, L., Nesbitt, R., Bloom, S., Rast, J.P., Berney, K., Arenas-Mena, C., Martinez, P., Lucas, S., Richardson, P.M., Davidson, E.H., Peterson, K.J., Hood, L., 2006. Unusual gene order and organization of the sea urchin hox cluster. J. Exp. Zool. B Mol. Dev. Evol. 306, 45–58.

Carroll, S.B., 1995. Homeotic genes and the evolution of arthropods and chordates. Nature 376, 479–485.

Coulier, F., Popovici, C., Villet, R., Birnbaum, D., 2000. MetaHox gene clusters. J. Exp. Zool. 288, 345–351.

Davidson, E.H., Erwin, D.H., 2006. Gene regulatory networks and the evolution of animal body plans. Science 311, 796–800.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Rust, A.G., Pan, Z., Schilstra, M.J., Clarke, P.J., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., Bolouri, H., 2002. A genomic regulatory network for development. Science 295, 1669–1678.

Dobias, S.L., Zhao, A.Z., Tan, H., Bell, J.R., Maxson, R., 1996. *SpHbox7*, a new Abd-B class homeobox gene from the sea urchin *Strongylocentrotus purpuratus*: insights into the evolution of hox gene expression and function. Dev. Dyn. 207, 450–460.

Dolecki, G.J., Wannakrairoj, S., Lum, R., Wang, G., Riley, H.D., Carlos, R., Wang, A., Humphreys, T., 1986. Stage-specific expression of a homeo box-containing gene in the non-segmented sea urchin embryo. EMBO J. 5, 925–930.

Dolle, P., Izpisua-Belmonte, J.C., Falkenstein, H., Renucci, A., Duboule, D., 1989. Coordinate expression of the murine Hox-5 complex homoeobox-containing genes during limb pattern formation. Nature 342, 767–772.

Duboule, D., 1994. Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. Dev. Suppl. 135–142.

Duboule, D., 1998. Vertebrate hox gene regulation: clustering and/or colinearity? Curr. Opin. Genet. Dev. 8, 514–518.

Fang, R., Olds, L.C., Sibley, E., 2006. Spatio-temporal patterns of intestine-specific transcription factor expression during postnatal mouse gut development. Gene Expr. Patterns 6, 426–432.

Ferrier, D.E., Holland, P.W., 2001. Sipunculan ParaHox genes. Evol. Dev. 3, 263–270.

Ferrier, D.E., Holland, P.W., 2002. *Ciona intestinalis* ParaHox genes: evolution of Hox/ParaHox cluster integrity, developmental mode, and temporal colinearity. Mol. Phylogenet. Evol. 24, 412–417.

Ferrier, D.E., Dewar, K., Cook, A., Chang, J.L., Hill-Force, A., Amemiya, C., 2005. The chordate ParaHox cluster. Curr. Biol. 15, R820–R822.

Frobius, A.C., Seaver, E.C., 2006. ParaHox gene expression in the polychaete annelid *Capitella* sp. I. Dev. Genes Evol. 216, 81–88.

Garcia-Fernandez, J., 2005. The genesis and evolution of homeobox gene clusters. Nat. Rev., Genet. 6, 881–892.

Kamm, K., Schierwater, B., in press. Ancient complexity of the non-Hox ANTP gene complement in the anthozoan *Nematostella vectensis*. Implications for the evolution of the ANTP superclass. J. Exp. Zoolog. B Mol. Dev. Evol. (electronic publication ahead of print).

Kauffman, S.A., Shymko, R.M., Trabert, K., 1978. Control of sequential compartment formation in *Drosophila*. Science 199, 259–270.

Kim, Y., Nirenberg, M., 1989. *Drosophila* NK-homeobox genes. Proc. Natl. Acad. Sci. U. S. A. 86, 7716–7720.

Lewis, E.B., 1978. A gene complex controlling segmentation in *Drosophila*. Nature 276, 565–570.

Long, S., Martinez, P., Chen, W.C., Thorndyke, M., Byrne, M., 2003. Evolution of echinoderms may not have required modification of the ancestral deuterostome HOX gene cluster: first report of PG4 and PG5 Hox orthologues in echinoderms. Dev. Genes Evol. 213, 573–576.

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

Martinez, P., Lee, J.C., Davidson, E.H., 1997. Complete sequence of *SpHox8* and its linkage in the single Hox gene cluster of *Strongylocentrotus purpuratus*. J. Mol. Evol. 44, 371–377.

Martinez, P., Rast, J.P., Arenas-Mena, C., Davidson, E.H., 1999. Organization of an echinoderm Hox gene cluster. Proc. Natl. Acad. Sci. U. S. A. 96, 1469–1474.

McGinnis, W., Krumlauf, R., 1992. Homeobox genes and axial patterning. Cell 68, 283–302.

Minokawa, T., Rast, J.P., Arenas-Mena, C., Franco, C.B., Davidson, E.H., 2004. Expression patterns of four different regulatory genes that function during sea urchin development. Gene Expr. Patterns 4, 449–456.

Morris, V.B., Byrne, M., 2005. Involvement of two Hox genes and Otx in echinoderm body-plan morphogenesis in the sea urchin *Holopneustes purpurescens*. J. Exp. Zool. B Mol. Dev. Evol. 304, 456–467.

Nam, J., Nei, M., 2005. Evolutionary change of the numbers of homeobox genes in bilateral animals. Mol. Biol. Evol. 22, 2386–2394.

Negre, B., Ranz, J.M., Casals, F., Caceres, M., Ruiz, A., 2003. A new split of the Hox gene complex in *Drosophila*: relocation and evolution of the gene labial. Mol. Biol. Evol. 20, 2042–2054.

Nemer, M., Rondinelli, E., Infante, D., Infante, A.A., 1991. Polyubiquitin RNA characteristics and conditional induction in sea urchin embryos. Dev. Biol. 145, 255–265.

Peterson, K.J., 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, S.L., Holland, P.W., 2000. Evidence for 14 homeobox gene clusters in human genome ancestry. Curr. Biol. 10, 1059–1062.

Rast, J.P., Amore, G., Calestani, C., Livi, C.B., Ransick, A., Davidson, E.H., 2000. Recovery of developmentally defined gene sets from high-density cDNA macroarrays. Dev. Biol. 228, 270–286.

Seo, H.C., Edvardsen, R.B., Maeland, A.D., Bjordal, M., Jensen, M.F., Hansen, A., Flaat, M., Weissenbach, J., Lehrach, H., Wincker, P., Reinhardt, R., Chourrout, D., 2004. Hox cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*. Nature 431, 67–71.

Sharkey, M., Graba, Y., Scott, M.P., 1997. Hox genes in evolution: protein surfaces and paralog groups. Trends Genet. 13, 145–151.

Simeone, A., Acampora, D., Arcioni, L., Andrews, P.W., Boncinelli, E., Mavilio, F., 1990. Sequential activation of HOX2 homeobox genes by retinoic acid in human embryonal carcinoma cells. Nature 346, 763–766.

Swofford, D.L., 2000. PAUP: Phylogenetic Analysis Using Parsimony and Other Methods. Sinauer, Sunderland, MA.

Walters, J.R., Howard, A., Rumble, H.E., Prathalingam, S.R., Shaw-Smith, C.J., Legon, S., 1997. Differences in expression of homeobox transcription factors in proximal and distal human small intestine. Gastroenterology 113, 472–477.

Wang, D.G., Kirchhamer, C.V., Britten, R.J., Davidson, E.H., 1995. *SpZ12-1*, a negative regulator required for spatial control of the territory-specific *CyIIIa* gene in the sea urchin embryo. Development 121, 1111–1122.

Yaguchi, S., Katow, H., 2003. Expression of tryptophan 5-hydroxylase gene during sea urchin neurogenesis and role of serotonergic nervous system in larval behavior. J. Comp. Neurol. 466, 219–229.