

# Lawrence Berkeley National Laboratory

## Lawrence Berkeley National Laboratory

### Title

Unusual Gene Order and Organization of the Sea Urchin Hox Cluster

### Permalink

<https://escholarship.org/uc/item/6rx5g7jv>

### Authors

Richardson, Paul M.  
Lucas, Susan  
Cameron, R. Andrew  
et al.

### Publication Date

2005-05-10

Peer reviewed

## Unusual Gene Order and Organization of the Sea Urchin Hox Cluster<sup>†</sup>

R. Andrew Cameron<sup>1\*</sup>, Lee Rowen<sup>2\*</sup>, Ryan Nesbitt<sup>2</sup>, Scott Bloom<sup>2</sup>, Jonathan P. Rast<sup>3</sup>, Kevin Berney<sup>1</sup>, Cesar Arenas-Mena<sup>4</sup>, Pedro Martinez<sup>5</sup>, Susan Lucas<sup>6</sup>, Paul M. Richardson<sup>6</sup>, Eric H. Davidson<sup>1</sup>, Kevin J. Peterson<sup>7</sup>, and Leroy Hood<sup>2</sup>

<sup>1</sup>Division of Biology and the Beckman Institute, California Institute of Technology, Pasadena, CA 91125; <sup>2</sup>The Institute for Systems Biology, Seattle, WA 98103-8904; <sup>3</sup>Sunnybrook and Women's College Health Sciences Centre, Toronto, Ontario, M4N 3M5 Canada; <sup>4</sup>Department of Biology, San Diego State University, San Diego, CA. 92182-4614; <sup>5</sup>Departament de Genètica, Universitat de Barcelona, 08028-Barcelona, Spain; <sup>6</sup>US Department of Energy Joint Genome Institute, Walnut Creek, CA 94598; <sup>7</sup>Department of Biological Sciences, Dartmouth College, Hanover, New Hampshire 03755

\* These authors contributed equally to this work.

26 pages, 0 tables, 6 figures

running title: unusual sea urchin hox gene order

\*Correspondence to:  
R. Andrew Cameron  
Beckman Institute 139-74  
California Institute of Technology  
1200 East California Blvd.  
Pasadena, CA 91125  
phone: 626-395-8421  
fax: 626-795-3382

keywords: genomes, synteny, evolution, *Strongylocentrotus purpuratus*

<sup>†</sup>Support NIH: HD05753; NSF IBN9604454; Beckman Institute, California Institute of Technology; DoE LLNL W-7405-Eng-48; DoE LBNL DE-AC03-76SF00098; and DoE LANL W-7405-ENG-36.

## Abstract

The highly consistent gene order and axial colinear expression patterns found in vertebrate hox gene clusters are less well conserved across the rest of bilaterians. We report the first deuterostome instance of an intact hox cluster with a unique gene order where the paralog groups are not expressed in a sequential manner. The finished sequence from BAC clones from the genome of the sea urchin, *Strongylocentrotus purpuratus*, reveals a gene order wherein the anterior genes (Hox1, Hox2 and Hox3) lie nearest the posterior genes in the cluster such that the most 3' gene is Hox5. (The gene order is : 5'-Hox1, 2, 3, 11/13c, 11/13b, 11/13a, 9/10, 8, 7, 6, 5 - 3'). The finished sequence result is corroborated by restriction mapping evidence and BAC-end scaffold analyses. Comparisons with a putative ancestral deuterostome Hox gene cluster suggest that the rearrangements leading to the sea urchin gene order were many and complex.

## Introduction

The first evidence for the importance of the gene order in the hox cluster came from genetic studies in *Drosophila melanogaster* where it was shown that the role of an individual hox gene in axial specification and its position in the chromosomal cluster were colinear (Lewis, 1978; Kauffman et al, 1978). This observation was soon confirmed in other bilaterian animals, especially vertebrates (reviewed in Martinez and Amemiya, 2002). Molecular studies then revealed a precise co-linear relationship between gene position in the hox cluster and expression patterns in embryonic structures (Hunt and Krumlauf, 1991). The idea that the nested expression patterns of hox genes control axial specification in embryonic structures soon followed (Kessel and Gruss, 1991). Thus hox proteins, expressed in an axial arrangement, activate downstream regulatory programs and thereby lead to differential specification along the axis of different body parts, for example, the hindbrain (Keynes and

Krumlauf, 1994). One of the simplest hypotheses that explains the conservation of gene order is the idea that adjacent hox genes share cis-regulatory elements and thus separating the genes would be deleterious. However, no one mechanism of this sort has emerged to account for the conservation across many taxa and recent evidence shows that bilaterian embryos use a variety of strategies (reviewed in Kmita and Duboule, 2003).

One or more clusters of hox genes is a trait that is uniquely shared across bilaterian genomes investigated to date (Adoutte et al, 2000; Balavoine et al, 2002). Vertebrates typically possess four clusters except teleost fishes which have even more. Across the bilaterians the precise number of hox genes and their position in the cluster probably varies more from taxa to taxa than has been indicated in the cluster structure of sequenced vertebrate genomes. Unfortunately for this question, most of the hox gene surveys done in a wide variety of animal groups have identified paralog group members based on sequence similarity (reviewed in Martinez and Amemiya, 2002) and thus do not reveal order. Gene order determinations require fully sequenced genome cluster regions or combinatorial mapping by hybridization of hox gene coding regions to large insert genomic clones such as BACs and PACs. Among echinoderm classes, one or more representatives have been surveyed for hox gene diversity, but only in the sea urchin has a mapping analysis been performed (Martinez et al, 1999). Here we report the complete genomic sequence of the *S. purpuratus* hox gene cluster and augment the original mapping study. We describe the unique gene order in the sea urchin hox cluster, other genomic features of the region and consider the evolutionary implications of these results.

## **Materials and Methods**

**Probes, libraries, and screening.** In order to screen libraries for BAC genomic clones, probes from several previous studies were used (Arenas-Mena et al., 1998, 2000; Martinez et al., 1999). Two

homeobox genes, SpHox7 and SpHox11/13b, are expressed in embryos and full length cDNAs are available (Dobias et al, 1996; Zhao et al, 1991). A full-length cDNA for SpHox8 was identified by a PCR strategy also (Martinez et al. 1997). From one of the PAC clones isolated in a previous study (Martinez et al., 1999) a Hox1 fragment was identified by degenerate PCR. All of the other probe sequences were first identified by degenerate PCR from genomic DNA or lambda genomic libraries. Probes that did not contain the homeodomain were selected wherever possible. The probes were tested to verify that they were all from single-copy genes using a conventional 3 restriction enzyme strategy, pulsed field electrophoresis and DNA genome blots. The final set of probes was used to screen a *S. purpuratus* large BAC library with an average insert length of 140 Kb (Cameron et al, 2000). PCR was used to confirm the presence of the probe sequence in the hybridizing BACs and to explore the presence of all the other known Hox genes. This combined hybridization and PCR analysis provided a group of BAC clones that includes 9 of the previously identified hox genes (Figure 2): *SpHox2*, *SpHox3*, *SpHox4/5*, *SpHox6*, *SpHox7*, *SpHox8*, *SpHox9/10*, *SpHox11/13a* and *SpHox11/13b*. Once ordered-and-oriented sequence was obtained from selected hox gene-containing BACs, matches to *S. purpuratus* BAC end sequences in GenBank were used to identify BACs that extended the tiling path to include SpHox1 and the neighboring genes on each side of the complete hox cluster (Figure 2). As a result of the search the ordered and oriented sequence of BAC 79A2 (AC131510) was also identified as part of the tiling path.

We have used the HUGO Gene Nomenclature Committee (<http://www.gene.ucl.ac.uk/nomenclature/>) designations for the individual hox genes. Groups of orthologous genes that arose from earlier duplications are referred to as paralogy groups and follow the vertebrate nomenclature (Scott, 1992).

**Sequencing and assembly.** *S. purpuratus* BAC clones were sequenced using standard shotgun procedures (Rowen et al, 1999; see also <http://www.jgi.doe.gov/sequencing/protocols/index.html>). At Joint Genome Institute, individual BAC DNA was isolated using Qiagen's Large-Construct isolation kit (Qiagen, Valencia, CA), and randomly sheared using a Hydroshear (GeneMachines, San Carlos, CA). Fragments were blunt-end repaired and inserts of 2-4kb were purified from an agarose gel and ligated into pUC18. After transformation, plasmid DNA was prepared from 1ml of cell culture using the Brinkman Perfect Prep robot or with magnetic bead prep (SPRI: <http://www.jgi.doe.gov/sequencing/protocols/index.html>). Detailed protocols are published in Detter et al (2002). At Institute for Systems Biology, BAC DNA was prepared on the AutoGen 740 according to the manufacturer's instructions and subjected to sonication to generate short fragments. Inserts of 2-4 kb were extracted from a low-melt agarose gel and ligated into pUC18. After transformation, plasmid DNA was prepared from 1 ml of cell culture using the Brinkman PerfectPrep robot and reagents according to the manufacturer's instructions. DNA was sequenced using Applied Biosystems Big Dye terminators and loaded onto either a 3700 or 3730 Capillary sequencer. Reads were assembled and edited using the Phred-Phrap-Consed suite of programs (Gordon et al, 1998). Gaps were filled and low-quality regions resolved by resequencing with Big dye primers or by oligonucleotide-directed sequencing. The consensus sequence was generated with about 15-20 reads/kb on average. The contiguous finished hox gene cluster, including annotation of the spans covered by the individual BAC clones from which the sequence was derived, is found in GenBank accession (to be filled in later).

**Restriction fragment analysis.** In order to verify that the BAC insert sequences were correctly cloned from genomic DNA, a restriction fragment analysis of the genomic region was performed. Restriction enzymes for which sites were sparsely distributed in the tract between the Hox11/13b and the Hox3 coding regions were used to digest genomic DNA from the same individual from which the

library was made. Partial digests of SgfI, SfiI, RsrII, SgrAI, and NotI were each separated on 1% agarose gels in a Biorad pulse field gel apparatus. DNA fragments were transferred to Hybond N membranes and hybridized with specific gene probes labeled with P<sup>32</sup>-CTP by random priming using a kit according to the manufacturer's protocol. Autoradiograms were exposed on Kodak film and scanned.

**Phylogenetic analysis.** Sequences were edited and aligned with MacVector 7.0. Seventy-eight amino acids (aa) (60 aa from the homeodomain, and the 6 N-terminal and 12 C-terminal aa on either side of the homeodomain where available) from 126 *Hox* genes were analyzed with PAUP v. 4.0b10 for Macintosh (Sinauer Associates, Sunderland, Massachusetts) using three *S. purpuratus* homeodomain sequences as outgroups (*SpEngrailed*, *SpEve* and *SpNK2.1*). Only relatively complete homeodomain encoding sequences were used except for *Am9* where an incomplete homeodomain was chosen specifically to test whether it is orthologous to the *SpHox11/13c* gene. 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

**Gene Assignments.** The finished sequence of the sea urchin Hox cluster affords the opportunity to determine more carefully the orthology relationship of the individual sea urchin hox genes to hox genes identified in other bilaterian species. Using cDNA alignments, PFAM searches on homeodomains and *ab initio* exon predictions in concert with searches against the NCBI nr database, we collected probable single exons containing a homeodomain for each of the 10 genes previously identified in the cluster and an additional gene described below. Alignments that included the homeodomains and some adjacent sequence were determined and a distance-based analysis (minimum evolution) was performed (Figure 1). This tree is consistent with previous orthology assignments (Martinez et al, 1997; Arenas-

Mena et al, 1998; Peterson, 2004). The *S. purpuratus* Hox1 and Hox2 genes cluster unambiguously with both protostome and deuterostome members of their respective paralogy groups. The Hox5 gene previously clustered with both Hox4 and Hox5 paralogy groups and no distinguishing features allowed a unique assignment. As shown by Peterson (2004) and confirmed here, this gene can now be unequivocally assigned to the Paralog Group 5. The central group genes (Hox6, Hox7, and Hox8) are not unambiguously separated although the *S. purpuratus* members aggregate in a large clade with all the other representatives of this class. There are insufficient sequence signatures to discriminate among the central group genes analyzed here. The new gene revealed computationally from our genomic sequence groups with the ambulacrarian (echinoderm + hemichordate) 11/13 genes, specifically with the asteroid posterior gene Am9, and is named Hox 11/13c. There is strong support for the monophyly of both the 11/13a and 11/13b-c groups, but no support for orthology between the 11/13c from the hemichordates and the 11/13c from the echinoderms. Because the urochordate, vertebrate and protostome posterior genes group separately from this Hox11/13 clade of the ambulacrarian deuterostomes, it is reasonable to suggest that this is an independent duplication of the posterior genes in the lineage leading to the echinoderms and hemichordates (Peterson, 2004).

**Gene Order.** The sequenced BAC clones were assembled into two supercontigs that overlap by 50 kb. One supercontig is from the maternal chromosome and the second from the paternal chromosome, both represented in the BAC library. The two supercontigs were merged to form a single sequence 936 Kb in length. The 9 previously defined homeodomains and two full-length cDNA sequences (SpHox7 and SpHox11/13b) uniquely aligned with the genomic sequence (Figure 2).

At the 3' end of the cluster the first hox gene encountered is the Hox 5 gene. Moving along the cluster in the 5' direction the genes are sequential up to the Hox11/13c gene (Figure 2). Following a 102 kb intergenic region the remaining three genes of the hox cluster are in reverse order: Hox3, Hox2,



Hox1. The ends of the hox cluster are marked by several genes whose position has been conserved. The Hox5 gene lies nearest to an acetylcholinesterase gene which is in turn adjacent on the 3' side to an alpha-2 collagen gene. Adjacent to Hox 1 at the 5' end of the cluster is an Eve orthologue previously identified as a component of the endomesoderm gene regulatory network (Davidson et al, 2002).

**Restriction fragment analysis.** In order to provide independent validation of the arrangement of hox genes in the sea urchin genome as revealed by the BAC sequences, a partial-digest restriction fragment analysis was performed using the same single animal DNA from which the BAC library was made (Figure 3). Taking the genomic sequence as a starting point, the sites for three restriction enzymes that cut rarely in the Hox genomic region, SgfI, RsrII, and SgrAI, were identified. Each of these enzymes is predicted to span a different portion of the region between Hox11/13a and Hox1 (Figure 3A). Three SgrAI sites included two predicted fragments 230 and 245 kb in length which included Hox11/13b, Hox11/13c and Hox3. All three of the probes used: Hox11/13b, Hox11/13c and Hox3 labeled two fragments of the predicted size (Figure 3B). In a like manner, RsrII is predicted to produce a 151kb fragment that includes Hox 11/13c and Hox3. The blot shows a doublet in RsrII digests at the correct size when probed with Hox11/13c and Hox3 but no bands in the Hox11/13b blot. The doublet is most likely due to an undiscovered polymorphism or a partial digest. The digests with SgfI should include Hox11/13b and Hox11/13c but not Hox3. The blot is consistent with that interpretation, viz. there are two bands at about 145 and 165 kb in both of the Hox11/13 blots. The SfiI and NotI digests were included on the blot but were not diagnostic for the features in question. The restriction fragments predicted from the finished sequence do indeed contain the expected genes and thus confirm the sequence derived from the BAC clones.

**BAC-end Scaffolds.** Because of the unusual gene order revealed in the finished sequence of the Hox cluster we sought further validation in addition to the restriction enzyme mapping described above

in order to verify the accuracy of the sequence. We used Blast to compare the finished hox cluster sequence in 100 kb increments against the GSS division of Genbank which contains a large number of *S. purpuratus* BAC-end sequences (Cameron et al, 2000). To identify matches to unique sequence, two iterations of masking out repeats from the hox sequence were performed, and then the masked sequences compared to GSS sequences again. To further guarantee the validity of the end-sequence matches, only those end reads with an EcoR1 site (allowing for one mispaired base) were retained. The EcoR1 restriction enzyme site was used for cloning the genomic DNA used in the library (Cameron et al, 2000). Along with end sequences corresponding to BACs that were completely sequenced, 25 additional matching BAC end pairs were identified. In several cases, only one end was retrieved from the blast matches but the other was found by a clone name search in GenBank and the match of this end to the hox cluster sequence verified. The 15 BAC end pairs that overlap the region of Hox11/13c and Hox3 are mapped onto that portion of the cluster sequence (Fig. 4). Since the BAC library from which the BAC-ends derive has an estimated genome coverage of 13X, the number of additional clones are consistent with that expected from the library and further substantiate the accuracy of the finished sequence.

**Genomic structure.** The hox cluster spans slightly over one half megabase (588) of genomic sequence (Figure 2). The intergenic distances vary from 8 kb between Hox 7 and Hox8 to over 100kb between Hox11/13c and Hox3. The other intergenic distances are all within 2-fold of the average figure of 30 kb determined from sequenced BACs and other genomic studies (Cameron et al, 2000). No other protein coding genes were identified within the cluster by routine sequence searches. Two partial retrotransposons are recognizable. A match to only a reverse transcriptase was found between the homeodomain exon matches for Hox 8 and 9/10. None of the other protein matches expected for the open reading frame of a retrotransposon were identified. However, both a reverse transcriptase and core

integrase match are found outside the cluster just past the even-skipped orthologue. In the latter case, the best match for the intervening reverse transcriptase is to a *Fugu* Line element belonging to the CR1 family (Poulter et al, 1999). The prevalence of genome-wide interspersed repeat sequences in the hox cluster was analyzed by comparison to a repeat collection that was derived from a self-search of BAC end sequences (Cameron et al, 2000). The distribution of these repeats in the cluster was compared to a concatenation of 10 BACs randomly taken from among the 40 ordered and oriented BAC sequences in Genbank. To assess the distribution of repeats in these two sets of sequence the repeats were compared to the sequences by Blast and the number of matches in each 5 Kb interval determined. The cluster had 100 empty intervals and the BAC sequence 42. Over the entire sequence the cluster averages 16 repeats per interval and the BACs 28. Thus, both the range of repeat frequencies and the overall distribution was similar between the two classes with a slightly lower figure in the hox cluster.

The alignment of the individual hox exons to the genomic sequence also reveal that the direction of transcription is not the same for all genes (Figure 2). The anterior genes (Hox1, Hox2, Hox3) are reversed with respect to their nearest neighbor Hox11/13c. Also Hox5 and Hox11/13b are reversed with respect to adjacent genes. The central group genes Hox6 through Hox11/13a are transcribed in the same direction.

**MiRNAs.** We scanned the assembled hox cluster for micro-RNAs using the rfam HMM models and Blast against the microRNA registry (Griffith-Jones, 2004) and confirmed the identification of a mir-10 ortholog near the Hox3 gene as originally shown by Tanzer and coworkers (2005).

## **DISCUSSION**

**Genomic Structure.** Finished genomic sequence of the purple sea urchin hox cluster occupies a region of 588 kb (Figure 2). When the adjacent non-Hox genes are included, the total region of finished

sequence spans 936 Kb. There are no intervening genes among the Hox cluster members. While at the time of this report there is no genome-wide repeat collection or RepeatMasker database for the sea urchin, a preliminary repeat collection was produced from an all-by-all search of 76,000 BAC end sequences (Cameron et al, 2000). A repeat found by this method is expected to occur at least 20 times in the whole genome since the BAC-end sequence database represents about 1/20 of the total genomic sequence. The sequence represented in repeats is 7% of the total. Analyzed with this repeat library the repeat content of the hox region, though a little sparse compared to the currently available BAC sequences, is not otherwise unusual. For example, there is no bias in the distribution of repeats as seen in the repeat-poor gnathostome genomic hox gene sequences (Fried et al, 2003). The remnants of reverse transcriptase coding sequence found in the cluster are far too divergent to have been active recently.

MicroRNAs may play an important role in regulating the expression of adjacent genes—hence we were interested in analyzing the hox cluster in this regard. In an evolutionary survey of microRNAs in Hox clusters, a sea urchin homolog of the mir-10 microRNA was identified from unassembled sea urchin genomic sequences (Tanzer et al, 2005). Another microRNA associated with vertebrate Hox clusters, mir-196, was not found in the sea urchin traces. We confirmed the identity of mir-10 in the sea urchin Hox cluster and fix its position between Hox11/13c and Hox3. We could not identify a homolog of mir-196 in the finished hox cluster sequence. This mir-10 microRNA was originally identified between Dfd (Hox4) and Scr (Hox5) in the *Drosophila melanogaster* HOX cluster and found to be conserved in mosquito and red flour beetle (Lagos-Quintana et al, 2001). Variants of this microRNA were found in zebrafish, pufferfish, mouse and human (Lagos-Quintana et al, 2003). One variant, mir-10b, lies between Hox4b and Hox5b in both mouse and human and the other, mir-10a, is found within intron 4 of Hoxd4 in the mouse genome and between Hoxd4 and Hoxd8 in the human genome. A recent

comparative analysis of the Hox3 and Hox4 genes in vertebrate taxa shows that an alternatively spliced exon of Hox3a in zebrafish is located only 10 Kb upstream of the adjacent Hox4a (Hadrys et al, 2004). These minor variations in position of the mir-10 microRNA do not reveal a clear ancestral position. However, the position of mir-10 near the Hox3 gene in the sea urchin which lacks a Hox4 gene is consistent with the general pattern seen elsewhere. The aforementioned mir-196 lies near the Hoxb8 gene in the mouse genome and it has been shown to direct the cleavage of Hoxb8 transcripts (Yekta et al, 2004). It is likely that conserved mir-10 similarly acts on the hox gene nearest to it which in the sea urchin is Hox3. Thus post-transcriptional mechanisms mediating expression patterns are likely conserved across the bilaterian taxa. Furthermore, the positional constraints for these non-coding RNAs to act on adjacent genes could contribute to the conservation of gene position in the cluster.

**Gene Identity.** The minimum-evolution tree (Figure 1) produced from 126 homeodomain sequences assigns the sea urchin Hox genes to canonical paralog groups consistent with previous reports (Martinez et al, 1999; Peterson, 2004). Correct phylogenetic assignment of the anterior Hox genes is crucial to our argument of cluster rearrangement (see below). The assignments of SpHox1, SpHox2, SpHox3 to their respective paralogy groups are unambiguous and supported by bootstrap values in excess of 70% (Fig. 1), reflecting the presence of paralogy-group-specific residues (Sharkey et al., 1997) in each individual SpHox gene. In addition, their integrity as a mini-cluster (see below) is further evidence for phylogenetic accuracy. With respect to SpHox5, although initially ambiguous (hence its original designation as SpHox4/5, Martinez et al., 1999), the discovery of both a Hox4 and a Hox5 in the hemichordate (Peterson, 2004) and asteroid echinoderm (Long et al., 2003) again allows for a clear and unambiguous assignment of this gene to the Hox5 paralogy group (Figure 1).

**Gene Order And Orientation.** The 11 hox genes display a unique order in which one end begins with the anterior genes (Hox1 and Hox2) and Hox3 genes and continues counting downward

from the posterior Hox 11/13c to the central group gene, Hox 5 (Figure 2). Thus the anterior genes lie next to the posterior ones but in reverse order. This arrangement is unique for sea urchins although few clusters have been mapped in this part of the deuterostome lineage. The closest comparisons among the deuterostomes are two urochordates: the ascidian, *Ciona intestinalis* (Dehal et al, 2001; Keys et al, 2005) and the appendicularian, *Oikopleura dioica* (Seo et al, 2004), and the cephalochordate, *Branchiostoma floridae* (Garcia-Fernandez and Holland, 1994; Ferrier et al, 2000). The HOX cluster genes of both urochordates are dispersed in a haphazard manner (Ikuta et al, 2004; Seo et al, 2004). The appendicularian genome has 9 recognizable hox genes all in separate linkage groups and no two genes could be linked by chromosome walking (Seo et al, 2004). From the ascidian genome draft sequence, 9 hox genes were identified on five different sequence scaffolds. Fluorescence in situ hybridization studies showed that 7 of the nine are on a single chromosome but not in the usual sequential order (Ikuta et al, 2004). In contrast, the cephalochordate cluster contains 14 paralog groups in a linear order from 1 to 14. Using both cephalochordate and protostomes as outgroups suggests that the loss of Hox4 and the rearrangements of the sea urchin cluster are conditions unique to this taxonomic group. Given that the rearrangement involves a break between Hox3 and Hox5 it can be suggested that the breakage is responsible for the loss of Hox4. That Hox4 is present in both asteroid echinoderms (Long et al., 2003) and hemichordates (Peterson, 2004), suggests that the rearrangement and loss happened in the *S. purpuratus* lineage after the split from asteroids. Ascertaining when it happened will require sequence and mapping data from holothurians and other echinoids.

**Syntenic And Rearrangement.** The syntenic relationships of the genes adjacent to members of the sea urchin cluster offers a curious comparison to the vertebrates (see Figure 6). There are three genes near the 5' end of the vertebrate hox clusters, the orthologs of two of these genes lie on one side of the sea urchin hox cluster and one on the other side. At the 5' end of the sea urchin cluster near the

Hox1 gene there is an Evx homolog about 64 kb away. At the 3' end of the sea urchin cluster, the Hox 5 gene is immediately flanked by an acetylcholinesterase gene about 50 Kb away. There is a more distant collagen transcription unit another 50 kb farther 5'. In both human and mouse genomes there is one member of an alpha fibrillar collagen gene family past the 3' end of each cluster (Bailey et al, 1997). The completed human sequence viewed in the NCBI Genome Viewer (<http://www.ncbi.nlm.nih.gov>) shows this gene to be about 60 Mb away from the HoxA cluster on chromosome 7. The sea urchin ortholog (Exposito et al, 1992) used in the mammalian synteny study as an outgroup (Bailey et al, 1997) is exactly the sequence we find about 100 kb away from the sea urchin Hox 5 paralog. In the human genomic region around the HoxA cluster, the acetylcholinesterase gene is about 7 Mb farther away in the 5' direction from the collagen gene. The completed genomic sequences displayed in the NCBI Genome Viewer shows the EVX genes to be the closest identified genes near the posterior end of the human HoxA and HoxD clusters. The human EVX1 gene lies quite close to the Hoxa13 gene, only 46 kb away. There is no paralog for this gene near the vertebrate HoxC or HoxD clusters. However, an Evx homolog is near Hox14 in the single cephalochordate hox cluster (Ferrier et al, 2001) and near a hox gene in cnidarians (Finnerty, 2001). The sea urchin Hox1 gene at the 5' end is flanked by an ortholog of Evx (Speve). A BAC sequence (AC131510) that had been previously characterized in an analysis of the gene encoding this developmentally regulated homeo-domain protein aligned with the hox cluster sequence placing Speve 75 kb 5' of the Hox1 gene. Thus the syntenic arrangement of these genes with respect to the hox cluster existed in the common ancestor of the chordates and echinoderms.

Using these syntenic relationships and deriving a putative ancestral cluster gene arrangement for the cephalochordate and vertebrate clusters (fig. 6A), one possible set of the steps in the re-arrangement leading to the sea urchin gene order and orientation can be postulated. The ancestral cluster was likely linear from Hox1 to some combination of posterior genes including the ancestor of Hox13. The

direction of transcription is from the most posterior gene toward the anterior ones and by convention the posterior gene is at the 5' end. The exact ancestral complement of posterior genes from Hox9 to Hox13 is ambiguous from the cephalochordate and vertebrate clusters but the expansion of the hemichordate and sea urchin 11/13 class is a clade-specific event (Peterson, 2004). It has previously been shown that the cephalochordate duplication leading to Hox14 is unique to that group (Garcia-Fernandez and Holland, 1994; Ferrier et al, 2000). The sea urchin hox cluster has the anterior three genes in a different position from the ancestor; Hox4 is lost; there is loss and duplication among the posterior genes; two of the sea urchin genes are transcribed from the opposite strand; and the genes adjacent to the cluster are in a different arrangement from the ancestor. Extensive translocations and inversions are necessary to reach the gene order and orientation detailed here. While it is difficult to know the exact order and extent of the various events it is clear that the changes were many and complex. Perhaps the strict adherence to a colinearity rule is no longer necessary in an animal that has secondarily gone from a bilaterian adult body plan to a pentamerous one.

**Colinearity.** The highly conserved colinear properties of HOX clusters and the gene regulatory mechanisms that may preserve them are a consistent feature of vertebrates and *Drosophila melanogaster* (Dolle and Duboule, 1989; Graham et al, 1989; see Duboule, 1998 for review) strongly suggesting that this is the primitive condition for bilaterians. Consistent with this suggestion is the presence of a single linear cluster in various arthropods including the red flour beetle, *Tribolium castaneum*, (Brown et al, 2002); the grasshopper, *Schistocerca gregaria*, (Ferrier and Akam, 1996) and the mosquito *Anopheles gambiae* (Powers et al, 2000; Devenport et al, 2002) (reviewed in Hughes and Kaufmann, 2002). Nonetheless, cluster evolution has occurred within the arthropods. For example, in the silkworm *Bombyx mori*, the Hox1 paralog (labial) is found some distance away from the rest of the cluster but on the same linkage group (Yasukochi et al, 2004). Among *Drosophila* species splits in the cluster are seen



at different locations (Von Allman et al, 1996; Lewis et al, 2003). Indeed, the labial gene of *Drosophila buzzati*, a member of the *Drosophila repleta* group, has become relocated near the posterior genes *abd-A* and *Abd-B* (Negre et al, 2003). Since all of these flies belong to the same genus and have similar axial morphology, it is unlikely that colinearity alone is necessary to maintain a similar axial specification system among them. Among other ecdysozoans, *Caenorhabditis elegans* has a total of 6 Hox genes that lie in two groups separated by 6 Mb (*C. elegans* Sequencing Consortium, 1998), but only three of these genes are required for normal embryogenesis (Van Auken et al, 2000). In the chordates both extremes are realized. The integrity of the Hox cluster is maintained in the cephalochordate (Garcia-Fernandez and Holland, 1994; Ferrier et al, 2000), but is dispersed in the appendicularian and ascidian urochordates despite the fact that the appendicularian expression patterns are axially ordered (Seo et al, 2004).

The unusual Hox gene order found in the sea urchin genome and the piecemeal pattern of expression in time and space break the rule of colinearity once again. Only two sea urchin genes are expressed in embryonic development, *SpHox7* (Angerer et al, 1989) and *SpHox11/13b* (Dobias et al, 1996). Their expression is dynamic, but by late embryogenesis the former is confined to subregions of the oral ectoderm and the latter to the vertex of the embryonic ectoderm (Angerer et al, 1989; Dobias et al, 1996). All of the Hox genes for which specific probes are available (*SpHox* 2, 3,5, 7, 8, 9/10, 11/13a and 11/13b) are expressed during the larval stages and in adult tissues (Arenas-Mena et al, 1998). In the course of larval development the most posterior genes (*SpHox*7, 8, 9/10, 11/13a, 11/13b) display a colinear expression pattern in the somatocoels, the paired posterior mesodermal structures that lie along the U-shaped gut and will become the adult perivisceral coeloms (Arenas-Mena et al, 2000; see Figure 5). The *Hox7* gene product is most anterior in the coelomic region that lies nearest the foregut and the territories of expression overlap in a consecutive series with *Hox11/13b* in the portion nearest the anus. These genes are also expressed in other tissues of the larva or presumptive adult but no such sequential

pattern is seen there (Arenas-Mena et al, 1998; Arenas-Mena et al, 2000). Thus, the original A/P axis orientation of the somatocoels is displayed in posterior hox gene expression and the emergence of five-fold radial symmetry must have been a later development (Peterson et al, 2000).

Using protostomes as an outgroup to indicate the direction of change within Hox cluster evolution, it seems that a single cluster of Hox genes in a large coherent cluster is primitive for Deuterostomia. These character states are maintained in amphioxus, but not in the urochordates thus far examined (see above). Within echinoids, a single large cluster is primitive, but the paralog group genes are no longer sequentially ordered. Vertebrates have maintained chromosomal colinearity, but have reduced the size of the cluster, as well as duplicated the cluster twofold. Vertebrates have also reduced the number of repeat sequences within the cluster as compared to the echinoid, but when this change occurred in the lineage is not clear. These phylogenetic comparisons suggest that the coordinated regulation of the compactly arranged Hox cluster genes of the vertebrates is a derived feature, and may not be primitive for either Deuterostomia or Bilateria. Data from other deuterostomes, especially hemichordates, will no doubt shed light on this issue.

#### ACKNOWLEDGMENTS

We acknowledge support by the National Institutes of Health Grant HD05753 (to E.H. D.), the National Science Foundation Grant IBN9604454 (to R.A.C.) and the Beckman Institute, California Institute of Technology (PI: E.H.D.). A portion of this work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program and the by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC03-76SF00098 and Los Alamos National Laboratory under contract No. W-7405-ENG-36.

## REFERENCES

- Adoutte A, Balavoine G, Lartillot N, Lespinet O, Prud'homme B, de Rosa R, 2000. The new animal phylogeny: Reliability and implications. *Proc Natl Acad Sci. USA* 97:4453-4456.
- Angerer LM, Dolecki GJ, Gagnon ML, 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 & Development* 3:370-383.
- Arenas-Mena, C, Cameron, RA, Davidson, EH, 2000. Spatial expression of Hox cluster genes in the ontogeny of a sea urchin. *Development* 127:4631-4643.
- Arenas-Mena, C, Martinez, P, Cameron, R. A. and Davidson, E. H, 1998. Expression of the *Hox* gene complex in the indirect development of a sea urchin. *Proc Natl Acad Sci USA* 95:13062-13067.
- Bailey WJ, Kim J, Wagner GP, Ruddle FH, 1997. Phylogenetic reconstruction of vertebrate hox cluster duplications. *Mol Biol Evol* 14:843-853.
- Balavoine G, de Rosa R, Adoutte, A, 2002. Hox clusters and bilaterian phylogeny. *Molecular Phylogenetics and Evolution* 24:366-373.
- Brown SJ, Fellers JP, Shippy TD, Richardson EA, Maxwell M, Stuart JJ, Denell, R, 2002. Sequence of the *Tribolium castaneum* Homeotic complex: The region corresponding to the *Drosophila melanogaster* Antennapedia complex. *Genetics* 160: 1067-1074.
- C. elegans* Sequencing Consortium 1998. *Science* 282:2012–2018.
- Cameron RA, Mahairas G, Rast JP, Martinez P, Biondi TR, Swartzell S, Wallace JC, Poustka AJ, Livingston BT, Wray GA, Etensohn CA, Lehrach H, Britten RJ, Davidson EH, Hood L, 2000. A sea urchin genome project: sequence scan, virtual map, and additional resources. *Proc Natl Acad Sci USA* 97:9514-9518.
- Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh C-H, Minokawa T, Amore G, Hinman V, Arenas-Mena C, Otim O, Brown CT, Livi CB, Lee P, Revilla R, Rust AG, Pan ZJ, Schilstra MJ, Clarke PJC, Arnone MI, Rowen L, Cameron RA, McClay DR, Hood L and Bolouri HA, 2002. Genomic regulatory network for development. *Science* 295:1669-1678.
- Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM, Harafuji N, Hastings KE, Ho I, Hotta K, Huang W, Kawashima T, Lemaire P, Martinez D, Meinertzhagen IA, Necula S, Nonaka M, Putnam N, Rash S, Saiga H, Satake M, Terry A, Yamada L, Wang HG, Awazu S, Azumi K, Boore J, Branno M, Chin-Bow S, DeSantis R, Doyle S, Francino P, Keys DN, Haga S, Hayashi H, Hino K, Imai KS, Inaba K, Kano S, Kobayashi K, Kobayashi M, Lee BI, Makabe KW, Manohar C, Matassi G, Medina M, Mochizuki Y, Mount S, Morishita T, Miura S, Nakayama A, Nishizaka S, Nomoto H, Ohta F, Oishi K, Rigoutsos I, Sano M, Sasaki A, Sasakura Y, Shoguchi E, Shin-i T, Spagnuolo A, Stainier D, Suzuki MM, Tassy O, Takatori N, Tokuoka M, Yagi K, Yoshizaki F, Wada S, Zhang C, Hyatt PD, Larimer F, Detter C, Doggett N, Glavina T, Hawkins T, Richardson P, Lucas S, Kohara Y, Levine M, Satoh N, Rokhsar DS, 2002. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* 298:2157-2167.

- Detter JC, Jett JM, Lucas SM, Dalin E, Arellano AR, Wang M, Nelson JR, Chapman J, Lou Y, Rokhsar D, Hawkins TL, Richardson PM, 2002. Isothermal Strand-Displacement Amplification Applications for High-Throughput Genomics. *Genomics* 80:691-698.
- Devenport MP, Blass C, Eggleston P, 2000. Characterization of the Hox gene cluster in the malaria vector mosquito, *Anopheles gambiae*. *Evol Dev* 2:326-339.
- Dobias SL, Zhao AZ, Tan H, Bell JR, 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. *Developmental Dynamics* 207: 450-460.
- Dolle P, Duboule D, 1989. 2 gene members of the murine hox-5 complex show regional and cell-type specific expression in developing limbs and gonads. *Embo Journal* 8:1507-1515.
- Duboule D, 1998. Vertebrate Hox gene regulation: clustering and/or colinearity?. *Curr Op Genet Dev* 8: 514-518.
- Exposito JY, Lessio M, Solursh M, Ramirez F. 1992. Sea-urchin collagen evolutionarily homologous to vertebrate pro- $\alpha$ -2(i) collagen. *J Biol Chem* 267:5559 -5562.
- Ferrier DEK, Akam M, 1996. Organization of the Hox gene cluster in the grasshopper, *Schistocerca gregaria*. *Proc Natl Acad Sci USA* 93:13024-13029.
- Ferrier D, Minguillon C, Cebrian C, Garcia-Fernandez J, 2001. Amphioxus *evx* genes: implications for the evolution of the midbrain-hindbrain boundary and the chordate tailbud. *Dev Biol* 237:270-281.
- Ferrier DEK, Minguillon C, Holland PWH, Garcia-Fernandez J, 2000. The amphioxus Hox cluster: deuterostome posterior flexibility and Hox14. *Evol Dev* 2:284-293.
- Finnerty JR, 2001. Cnidarians reveal intermediate stages in the evolution of hox clusters and axial complexity. *Amer Zool* 41:608-620.
- Fried C; Prohaska SJ; Stadler PF. 2003. Exclusion of repetitive DNA elements from gnathostome hox clusters. *J Exp Zool. (Mol Dev Evol)* 302B:165-173.
- Garcia-Fernandez J, and Holland P, 1994. Archetypal organization of the amphioxus Hox gene cluster. *Nature* 370:563-566.
- Gordon D, Abajian C, and Green P, 1998. Consed: a graphical tool for sequence finishing. *Genome Res* 8:195-202.
- Graham A, Papalopulu N, Krumlauf R, 1989 The murine and drosophila homeobox gene complexes have common features of organization and expression. *Cell* 57:367-378.
- Griffith-Jones S, 2004. The microRNA Registry. *Nucl Acids Res* 32: D109-111.
- Hadrys T, Prince V, Hunter M, Baker R, Rinkwitz S, 2004 Comparative genomic analysis of vertebrate Hox3 and Hox4 genes. *J Exp Zool Mol Dev Evol* 302:147-164.
- Hughes CL, Kaufman TC. 2002. Hox genes and the evolution of the arthropod body plan. *Evol Dev* 4:459-499.
- Hunt P, Krumlauf R, 1991. Deciphering the hox code - clues to patterning branchial regions of the head. *Cell* 66:1075-1078.

- Ikuta T, Yoshida N, Satoh N, Saiga H, 2004. *Ciona intestinalis* Hox gene cluster: Its dispersed structure and residual colinear expression in development. *Proc Natl Acad Sci USA* 101:15118-15123.
- Kauffman SA, Shymko RM, Trabert K, 1978. Control of sequential compartment formation in *Drosophila*. *Science* 199:259–270.
- Kessel M, Gruss P, 1991. Homeotic transformations of murine vertebrae and concomitant alteration of hox codes induced by retinoic acid. *Cell* 67: 89-104.
- Keys DN, Lee B, Di Gregorio A, Harafuji N, Detter JC, Wang M, Kahsai O, Ahn S, Zhang C, Doyle SA, Satoh N, Satou Y, Saiga H, Christian AT, Rokhsar DS, Hawkins TL, Levine, M, Richardson PM. 2005. A saturation screen for cis-acting regulatory DNA in the Hox genes of *Ciona intestinalis*. *PNAS USA* 102:679–683.
- Keynes R, Krumlauf R, 1994. Hox genes and regionalization of the nervous-system. *Ann Rev Neuroscience* 17:109-132.
- Kmita M, Duboule D, 2003. Organizing axes in time and space; 25 years of colinear thinking. *Science* 301:331-333.
- Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. 2001. Identification of novel genes coding for small expressed RNAs. *Science*. 294:853-858.
- Lagos-Quintana M, Rauhut R, Meyer J, Borkhardt A, Tuschl T, 2003. New microRNAs from mouse and human. *RNA* 9:175-179.
- Lewis EB, Pfeiffer BD, Mathog DR, Celniker SE, 2003. Evolution of the homeobox complex in the Diptera. *Curr Biol* 13:R587-R588.
- Lewis EB, 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 CT, 2002. Genomics of the HOX gene cluster. *Comp Biochem Physiol Biochem Mol Biol* 133:571-580.
- Martinez P, Lee JC, Davidson EH, 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 JP, Arenas-Mena C, Davidson EH, 1999. Organization of an echinoderm *Hox* gene cluster. *Proc Natl Acad Sci USA* 96:1469-1474.
- Negre B, Ranz JM, 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.
- Peterson KJ, Arenas-Mena C, Davidson, EH, 2000. The A/P axis in echinoderm ontogeny and evolution: Evidence from fossils and molecules. *Evo. Dev.* 2:93-101.
- Peterson, K.J. (2004). Isolation of *Hox* and *Parahox* genes in the hemichordate *Ptychodera flava*: Insights into the evolution of deuterostome *Hox* genes. *Mol Phylog Evol* 31:1208-1215.
- Poulter R, Butler M, Ormandy J, 1999. A LINE element from the pufferfish (fugu) *Fugu rubripes* which shows similarity to the CR1 family of non-LTR retrotransposons. *Gene* 227:169-179.

- Powers TP, Hogan J, Ke ZX, Dymbrowski K, Wang XL, Collins FH, Kaufman TC Characterization of the Hox cluster from the mosquito *Anopheles gambiae* (Diptera : culicidae). *Evol Dev* 2:311-325.
- Rowen L, Lasky S, Hood L, 1999. Deciphering genomes through automated large-scale sequencing. *Methods in Microbiology* 28:155-192.
- Scott MP, 1992. Vertebrate homeobox gene nomenclature. *Cell* 71:551-553.
- Seo H-C, Edvardsen RB, Maeland AD, Bjordal M, Jensen MF, Hansen A, Flaas 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 MP, 1997. *Hox* genes in evolution: protein surfaces and paralogy groups. *Trends Genet* 13:145-151.
- Tanzer A, Amemiya CT, Kim CB, Stadler PF, 2005. Evolution of microRNAs located within Hox gene clusters. *J Exp Zool Mol Dev Evol* 304:75-85.
- Van Auken K, Weaver DC, Edgar LG, Wood WB, 2000. Special Feature: *Caenorhabditis elegans* embryonic axial patterning requires two recently discovered posterior-group Hox genes *PNAS USA* 97:4499-4503.
- VonAllmen G, Hogga I, Spierer A, Karch F, Bender W, Gyurkovics H, Lewis E, 1996. Splits in fruitfly Hox gene complexes *Nature* 380:116-116.
- Yasukochi Y, Ashakumary LA, Wu CC, Yoshido A, Nohata J, Mita K, Sahara K, 2004. Organization of the Hox gene cluster of the silkworm, *Bombyx mori*: a split of the Hox cluster in a non-*Drosophila* insect. *Dev Genes Evol* 214:606-614.
- Yekta S, Shih IH, Bartel DP, 2004. MicroRNA-directed cleavage of HOXB8 mRNA. *Science* 304:594-596.
- Zhao AZ, Vansant G, Bell J, Humphreys T, Maxson R, 1991. Activation of the L1 late h2b histone gene in blastula-stage sea-urchin embryos by an Antennapedia-class homeoprotein. *Mech Dev* 34:21-28.

FIGURES

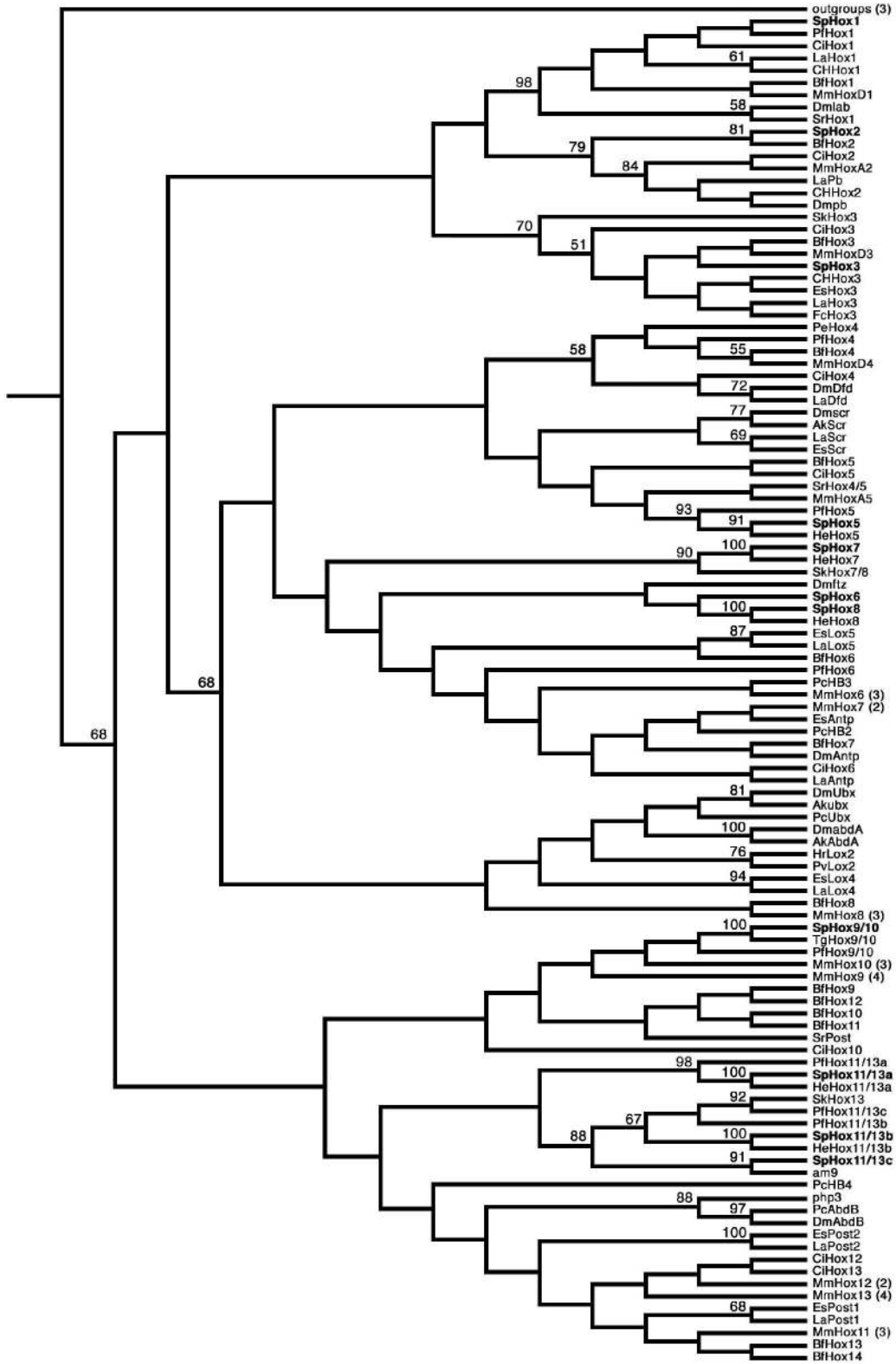


Fig. 1. The minimum-evolution tree for 126 *Hox* genes using three *S. purpuratus* homeodomain sequences as outgroups (score = 17.12807). The *S. purpuratus* genes are shown in bold. Bootstrap values > 50% are shown at the respective nodes. For simplicity, many of the vertebrate paralogues are not shown; instead the number of genes analyzed is given in parentheses behind the name. Taxonomic abbreviations are as follows: Ak – *Acanthokara kaputensis* (onychophoran); Am – *Asterina minor* (starfish); Bf – *Branchiostoma floridae* (amphioxus); CH – *Chaetopterus sp.*; Ci – *Ciona intestinalis* (ascidian); Dm – *Drosophila melanogaster* (fruit fly); Es – *Euprymna scolopes* (squid); Fc - *Folsomia candida* (collembolan insect); He – *Heliocidaris erythrogramma* (sea urchin); Hr – *Helobdella robusta* (leech); La – *Lingula anatina* (brachiopod); Mm – *Mus musculus* (mouse); Pc – *Priapulid caudatus* (priapulid worm); Pe – *Patiriella exigua* (starfish); Pf – *Ptychodera flava* (hemichordate); Pv – *Patella vulgata* (gastropod mollusc); Sk – *Saccoglossus kowalevskii* (hemichordate); Sp – *Strongylocentrotus purpuratus* (sea urchin); Sr – *Symsagittifera roscoffensis* (acoel flatworm); Tc – *Tribolium castaneum* (flour beetle); Tg - *Tripneustes gratilla* (sea urchin); php-3 is from *C. elegans*.

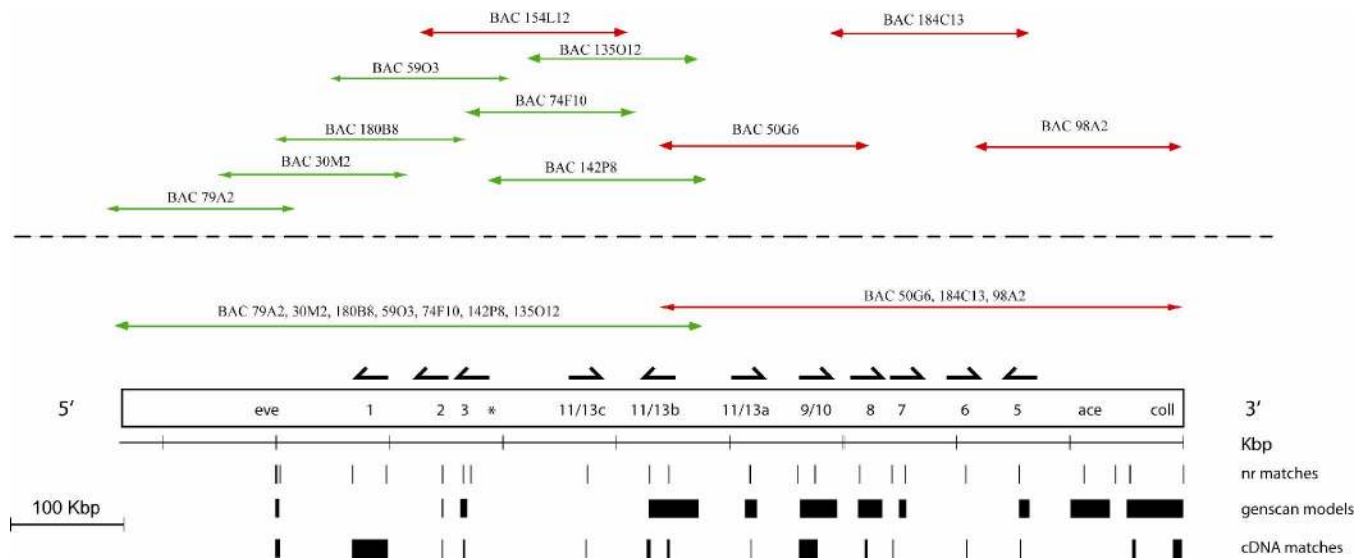


Figure 2. Gene order and BAC overlap in the sea urchin *hox* cluster. Above the dotted line are displayed the position of the individual BAC inserts that were sequenced in order to assemble the finished *Hox* cluster sequence. The lines are colored green or red to represent the two chromosomal alleles. Below the dotted line is shown the position of the two supercontigs, one from each allele represented in the library. Below the contigs, are the gene names and a line represents the assembled sequence. The direction of transcription is indicated above the boxed genes. The conserved miRNA, mir-10, is indicated with an asterisk. At the bottom is shown the various alignments to transcribed sequence. The alignments are of three categories: 1) sequences that match sequences in the non-redundant protein database from GenBank, nr matches (nr); 2) computed gene models by Genscan (Bureg and Karlin, 1997) from the genomic sequence that were subsequently compared to nr; 3) alignments to the genomic sequence by Blastn and Sim4 of the identified cDNA sequences ( see text). The cluster orientation (5' and 3') follows the usual convention.



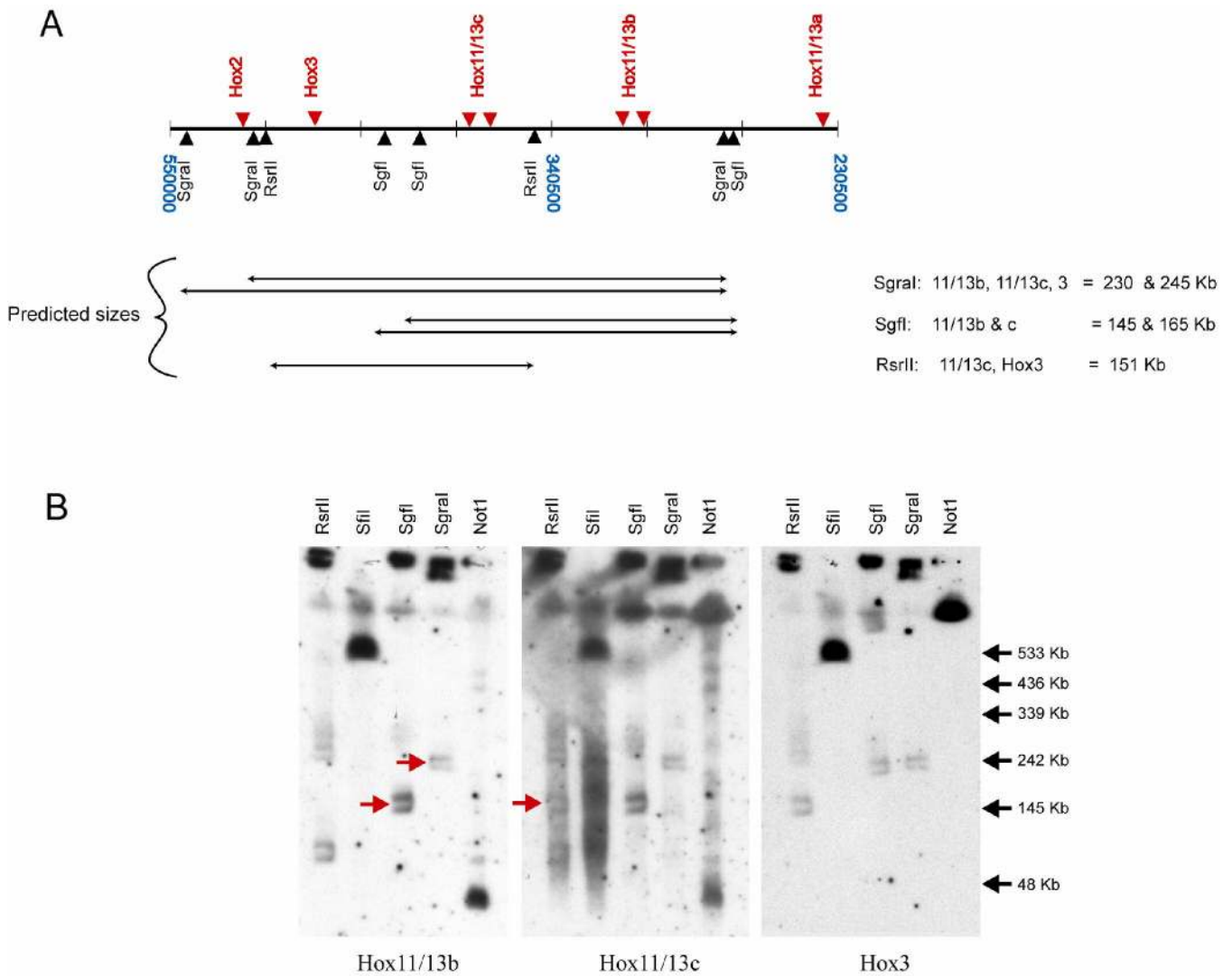


Figure 3. The restriction digests of genomic DNA that confirm the hox gene order. A. The predicted sites for rare cutting enzymes in the region of the cluster between the posterior class genes and the anterior class genes are displayed along with the predicted sizes of fragments. B. DNA gel blot hybridizations of a pulse-field gel separation of genomic DNA digested with the enzymes indicated across the top of the autoradiogram. The sizes for the markers used on this gel are indicated at the right. The red arrows indicate informative bands from the digests with SgfI, SgrI, and RsrII.

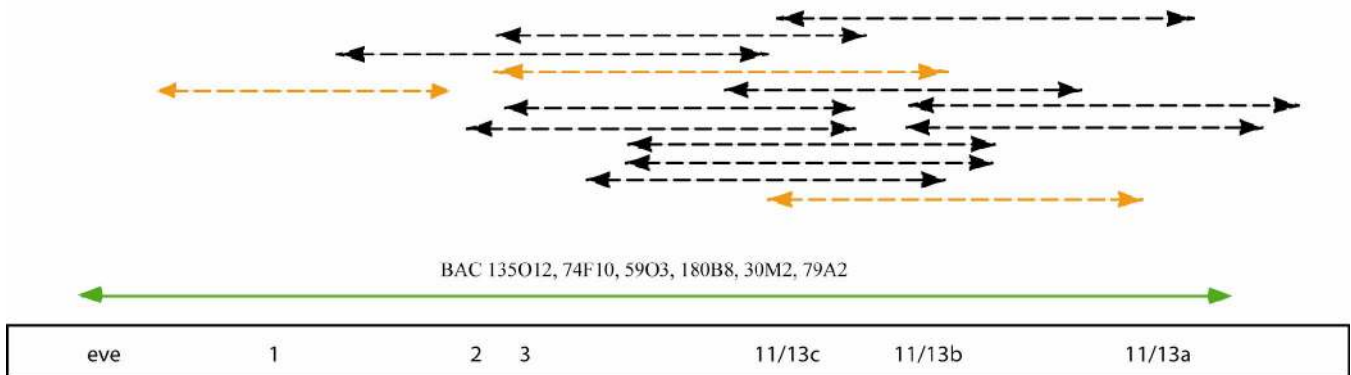


Figure 4. The position of the BAC-end sequences that lie specifically in region of probable re-arrangement (dotted box) are displayed. Those BACs for which both ends were matched by Blast alignments are shown in black while those for which one end was found by name are shown in orange.

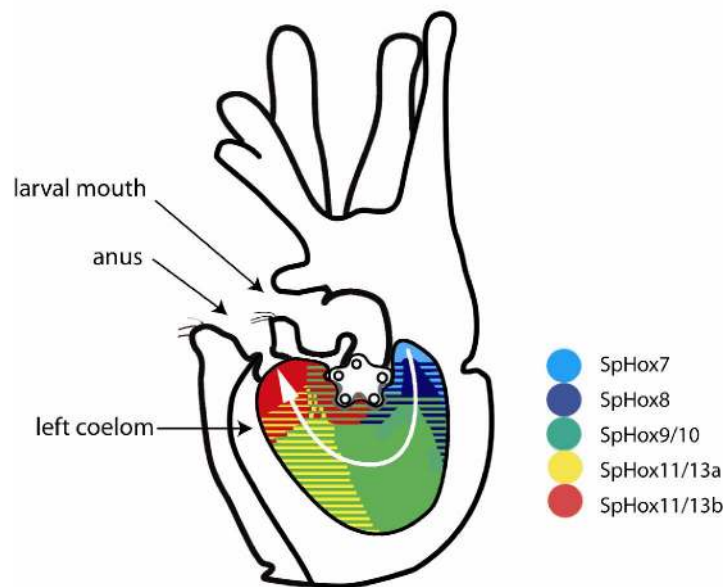


Figure 5. Diagram of posterior hox gene expression in the larval coelom (somatocoel). A late larval stage is viewed from the left side. The larval arms protrude upward and the mouth and anus are below the arms. The gut assumes a U-shape and the coelom lies against the stomach which is obscured in this diagram. The territories of hox gene expression which are depicted in colors (see legend) lie in a consecutive series along the gut from the mouth region to the anus region (redrawn from Arenas-Mena et al, 2002).

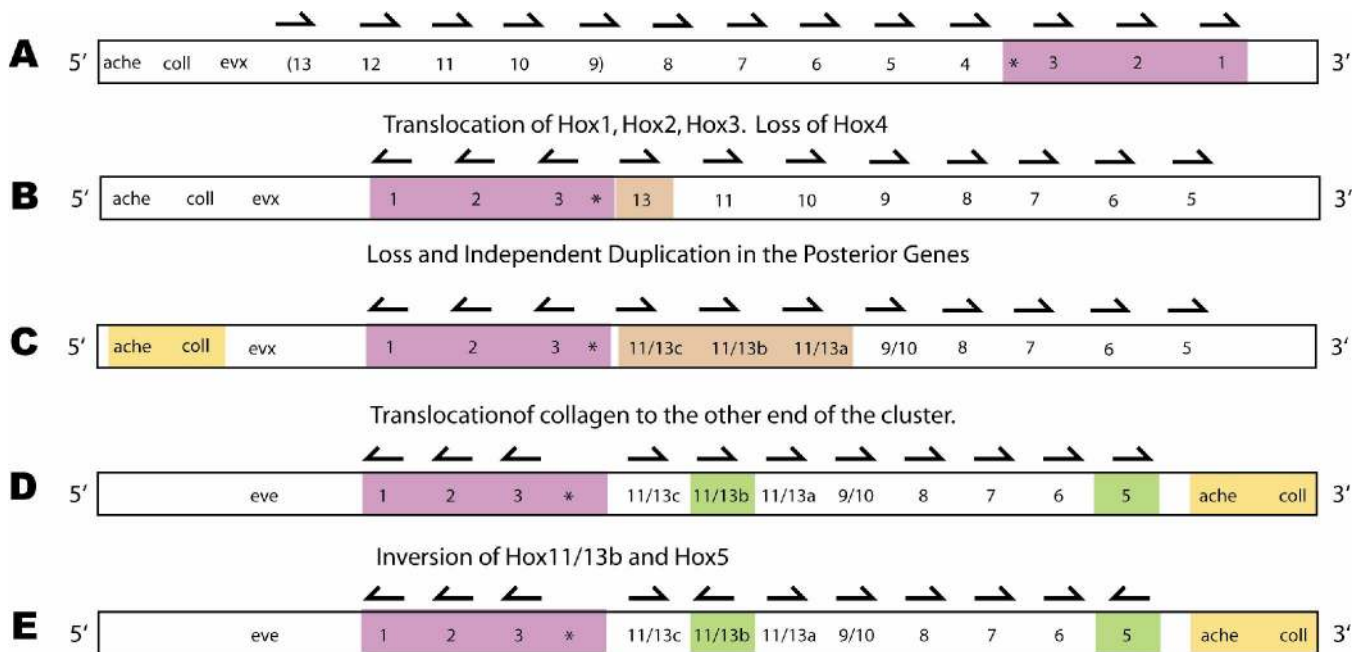


Figure 6. One of several equally parsimonious schemes for the rearrangement of a putative ancestral hox cluster to produce the order and orientation of genes seen in the sea urchin hox cluster. The gene names are listed within a long box denoting the sequence. Half-arrows above each gene show the direction of transcription. None of the intergenic distances are to scale. A. The putative ancestral consensus cluster derived from cephalochordate and vertebrate clusters (see text). The exact number and assignment of the posterior genes is not recoverable, hence they are within parentheses. B. The arrangement resulting from translocation and inversion of Hox1,2,3 to the opposite end of the cluster and the loss of Hox4. C. The further rearrangement emerging from losses and duplications of the most posterior genes. D. The translocation without inversion of collagen and acetylcholinesterase to the end of the cluster near Hox5. E. The in-place inversion of Hox11/13b and Hox5 to yield the orientation found from the finished sea urchin genomic sequence.

## **DISCLAIMER**

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.