Gene loss and gain in the evolution of the vertebrates

Frank H. Ruddle[1,4,*], Kevin L. Bentley[1], Michael T. Murtha[2] and Neil Risch[3,4]

[1]Department of Biology, Yale University, New Haven, CT, USA
[2]CuraGen Corporation, Branford CT 06405, USA
[3]Department of Epidemiology and Public Health, Yale University Medical School, New Haven, CT, USA
[4]Department of Genetics, Yale University Medical School, New Haven CT, USA

*Author for correspondence

SUMMARY

Homeobox cluster genes (Hox genes) are highly conserved and can be usefully employed to study phyletic relationships and the process of evolution itself. A phylogenetic survey of Hox genes shows an increase in gene number in some more recently evolved forms, particularly in vertebrates. The gene increase has occurred through a two-step process involving first, gene expansion to form a cluster, and second, cluster duplication to form multiple clusters. We also describe data that suggests that non-Hox genes may be preferentially associated with the Hox clusters and raise the possibility that this association may have an adaptive biological function. Hox gene loss may also play a role in evolution. Hox gene loss is well substantiated in the vertebrates, and we identify additional possible instances of gene loss in the echinoderms and urochordates based on PCR surveys. We point out the possible adaptive role of gene loss in evolution, and urge the extension of gene mapping studies to relevant species as a means of its substantiation.

Key words: homeobox, echinoderms, ascidians, gene families, genome duplication

INTRODUCTION

The homeobox system of genes is generally recognized as having useful attributes for the study of evolution (Shashikant et al., 1991; Kappen et al., 1993). Most prominent among these is its high level of conservation exemplified in the homeobox sequences and the structural organization of the Hox gene clusters. These properties make it possible to identify sequence motifs, genes, and gene clusters that are homologous, and thus can be compared both quantitatively and qualitatively with confidence over a broad spectrum of metazoans. That the homeobox genes play a fundamental role in metazoan development also suggests that they may themselves be important to the evolutionary process.

In a recent review (Ruddle et al., 1994), we have shown that homeobox genes have been reported for all the major phyla, this is also true for the clustered Hox genes with the exception of the sponges (Seimiya et al., 1992). Hox gene clusters have been directly demonstrated in *Caenorhabditis elegans* (Bürglin et al., 1991; Bürglin and Ruvkun, 1993), *Tribolium castaneum* (Beeman, 1987), *Drosophila pseudoobscura* (Randazzo et al., 1993), *D. melanogaster* (Lewis, 1978; Kaufman et al., 1990), *Branchiostoma floridae* (Holland et al., this volume; Pendleton et al., 1993), *Petromyzon marinus* (Pendleton et al., 1993), and all jawed vertebrates so far examined. Moreover, good evidence has been presented for Hox cluster duplication giving rise to four clusters, each on a different chromosome, in all amniotes (Kappen et al., 1989; Schughart et al., 1989).

In addition to the clustered Hox genes there exist a number of related homeobox genes which are divergent with respect to the homeobox and other features (Kappen et al., 1993). These fall into a number of groups based on similarity, as for example the *Paired*, *Caudal*, and *Distal-less* type genes. We will refer to these genes as non-clustered or diverged homeobox genes. The homeobox genes have been shown to undergo duplication by both *cis* (laterally within a chromosome) or *trans* (chromosome duplication within a genome) processes (Kappen and Ruddle, 1993). *Cis* duplication can arise by unequal crossing over and *trans* duplication by polyploidization, although other mechanisms are also possible.

Ohno has suggested that gene duplication by polyploidy serves an important role in evolution (Ohno, 1970). It is argued that developmentally relevant genes become integrated into developmental pathways that are hierarchical and highly interdependent, and thus they cannot readily mutate or take on new functions without disrupting the overall developmental plan. Gene duplication provides a way around this impasse by the retention of old developmental interrelations and the incorporation of newly duplicated genes into new pathways and relationships. The genomic and functional conservation of the homeobox system is consistent with this view (Ruddle et al., 1994). It is interesting that gene duplication is often cited as being adaptive, because it introduces genetic

redundancy into developmental systems. However, as viewed here, redundancy might simply be a consequence of developmental conservatism. Developmental genes and particularly homeotic genes are interactive transcriptionally and have been shown to have the properties of a combinatorial system (Wagner, 1994). The insertion of new genes into a network of genes can be expected to introduce new degrees of freedom, and thereby multiple possible avenues of evolutionary divergence. In this respect, the increase in Hox clusters and the duplication of many non-clustered homeobox genes in the forerunners of the vertebrates can be postulated to have had a profound effect on their evolution and possibly to have contributed to vertebrate radiation (Gould and Eldredge, 1993).

Previous studies have provided support for the idea that the vertebrate Hox cluster gene family has arisen by means of a two step process: firstly, the expansion of the cluster by lateral gene duplication, and secondly the duplication of the clusters by a large domain duplication events, as for example, chromosome or genome duplication (Ruddle, 1989; Kappen et al., 1989; Schughart et al., 1989). Sequence comparisons of the homeobox domain indicate a relatedness between paralogy groups 1-3, 4-8, and 9-13*. These groups have been termed anterior, medial, and posterior, respectively, based on their patterns of expression along the anterior/posterior axis (Kappen et al., 1989; Schubert et al., 1993; Ruddle et al., 1994) . These relationships have suggested that these three groups of genes have arisen from an ancestral cluster of three genes (Schubert et al., 1993).

In this report, we will discuss two systems that relate to chordate evolution and gene duplication. One deals with the amplification of non-homeobox genes which are in linkage with the Hox clusters. The second deals with the possible loss of Hox genes in echinoderms and tunicates.

PARALOGOUS GENES IN LINKAGE WITH THE Hox GENE CLUSTERS

An examination of genes in the vicinity of the Hox clusters shows that many are paralogous and map to two or more of the four Hox gene cluster chromosomes (Rabin, et al., 1986; Ferguson-Smith and Ruddle, 1988; Schughart, et al., 1988; Schughart, et al., 1989; Ruddle, 1989; Craig and Craig, 1991, 1992; Hart et al., 1992; Lundin; 1993; Bentley et al., 1993). In order to study these genes more systematically, we have tabulated all the mouse genes that are members of gene families and of which at least one member maps to the chromosomes bearing Hox gene clusters, namely chromosomes 2, 6, 11 and 15 (Silver, 1993; Siracusa and Abbot, 1993; Moore and Elliott, 1993; Buchberg and Camper, 1993; Mock et al., 1993). The Hox gene clusters themselves are not included in the sample to avoid bias. The identification of gene families is based on sequence similarity. Seventy-four families with a total of 323 genes were identified using these criteria. A representative sample of these genes including 30 families and 203 genes is shown in Fig. 1. A statistical analysis was used to test whether there was excess clustering of members of the

gene families on these four chromosomes†. The hypothesis of no excess clustering could be rejected at the 0.01 level of confidence. Four other chromosomes selected on the basis of size similarity to the Hox cluster-bearing chromosomes and number of mapped loci involved, namely mouse chromosomes 4, 9, 12, and 16, were subjected to the same analysis. In this instance the null hypothesis could be not rejected ($P = 0.11$).

The Hox gene clusters are estimated to have undergone duplication minimally 350 million years ago (Kappen et al., 1989). This figure may represent a gross underestimate since Forey and Janvier (1993) have determined the divergence date between lamprey and gnathostomes to be 435 million years. Sufficient time has elapsed since the duplication event to randomize genes throughout the genome. We base this supposition on the randomization of linkage relationships (demonstrated) between the mouse and human genomes over a period of approximately 100 million years (Nadeau, 1989, 1991). However, our analysis indicates a proclivity of genes linked to the clusters to retain their primordial linkage relationships. This relationship is all the more striking when one limits consideration to genes closely linked to the Hox gene clusters. In an extension of this study, we confined our analysis to genes mapping within approximately 30 centiMorgans (cM) centered on the Hox clusters. This distance represents approximately one half of chromosome 15, the smallest of the four chromosomes bearing Hox clusters. Paralogous genes linked within 15 cM to each side of the clusters showed a highly significant association with the clusters ($P = 2\times10^{-5}$). The probability score for paralogous genes outside this region on the same chromosomes was not significant ($P = 0.18$). Hence, the excess clustering of gene families initially observed for chromosomes 2, 6, 11, and 15 is due to specific clustering around the Hox gene complexes.

A possible explanation for these findings is that the linkage of genes to the Hox clusters is a simple structural remnant of the ancestral linkage pattern prior to cluster duplication. A second explanation is that the linkage of paralogous genes is conserved, because it serves an important biological function. This adaptive point of view is strengthened by the fact that many of the genes in linkage to the Hox cluster genes also serve a developmental function, such as growth factors, receptors, members of signaling pathways, and structural proteins having a developmental role such as the cytokeratins and collagens (see Fig. 1). It is also of interest and of possible significance that genes bearing a sequence or functional similarity to the mammalian genes in linkage to the Hox clusters

---

†First, we calculated the percentage of total mapped loci that fall on each chromosome. For example, the percentages for chromosomes 2, 6, 11, and 15 are 7.4%, 4.6%, 6.3% and 3.8%, respectively. For a gene family, a hit on a particular chromosome is defined as at least one member of the family mapping to that chromosome. For each gene family, there can be a total of 1 to 4 hits, depending on the number of chromosomes containing hits. There are four single hit possibilities (one for each chromosome), six possible two-hit combinations, four possible three-hit combinations, and one four-hit combination. The probability of each of these outcomes can be calculated directly from the percentages give above. We note that this probability needs to be corrected for the fact that the family must have contained at least one hit to be ascertained; hence, each multi-hit outcome probability is divided by the probability of at least one hit. This correction is similar, in spirit, to that used in segregation analysis for human recessive diseases, where families are ascertained through at least one affected child (Elandt-Johnson, 1971).

From these probabilities, we then calculate the expected number of single, double, triple, and quadruple hits, and compare these with the observed numbers. From the probability distribution for number of hits, we calculate the exact probability of obtaining the actual observed number of hits or greater using simulation; it is these P values that we report. A significant excess of hits over expected indicates clustering of gene families.

---

*Throughout this article paralogy is defined generally as homologies within a genome, but also more specifically to *trans*-homologies between chromosomes.

Paralogous Relationships of Genes Linked to Murine Hox Clusters

| Hoxd |
| Hoxa |
| Hoxb |
| Hoxc |

| Acra |
| Acra-4 |
| Acrb |
| 7 genes |

| Actc |
| Actg |
| Act-3 |
| Act-4 |
| Acta2-rs1 |
| 6 genes |

| Apoh |
| 4 genes |

| Brp-13 |
| Brp-1 |
| Brp-8 |
| 2 genes |

| Cola-2 |
| Cola-1 |
| Col2a-1 |
| 12 genes |

| Csfg |
| Csfgm |
| 1 gene |

| Dlx-1 |
| Dlx-2 |

| Erbb |
| Erbb-2 |
| 1 gene |

| Evx-2 |
| Evx-1 |

| Glut-3 |
| Glut-4 |
| 2 genes |

| Il-1a |
| Il-1b |
| Il-3 |
| Il-4 |
| Il-5 |
| Il-13 |
| 6 genes |

| Il-2ra |
| Il-5r |
| Il-2rb |
| Il-3rb |
| Il-7r |
| 4 genes |

| Ins-1 |
| Ins-3 |
| 1 genes |

| Itga4 |
| Itga6 |
| Itgav |
| Itga3 |
| Itga2b |
| Itga5 |
| 2 genes |

| Itgb6 |
| Itgb3 |
| Itgb4 |
| Itgb7 |
| 1 gene |

| Kras-3 |
| Kras-2 |
| 1 gene |

| Krt-1 |
| Krt-2 |

| Myhs |
| Myhs-e |
| Myhs-f2 |
| Myhs-p |
| Myh-9 |
| 2 genes |

| Pax-1 |
| Pax-6 |
| Pax-8 |
| Pax-4 |
| 4 genes |

| Rara |
| Rarg |
| 1 gene |

| Rbtn-2 |
| Rbtn-3 |
| 1 gene |

| Rpn-1 |
| Rpn-2 |

| Rxra |
| 2 genes |

| Scn1a |
| Scn2a |
| Scn3a |
| Scn4a |
| 2 genes |

| Tcrb |
| 4 genes |

| Top1 |
| Top2 |

| Wa-1 |
| Wa-2 |

| Wnt-2 |
| Wnt-3 |
| Wnt-1 |

Key
| Chr. 2 |
| Chr. 6 |
| Chr. 11 |
| Chr. 15 |
| Other |

Fig. 1. Paralogous relationships of genes linked to murine Hox clusters. Data used in preparing this figure were derived from The Encyclopedia of the Mouse Genome accessed at Jackson Laboratory and sources listed in the text. This table is a representative sampling of the multi-gene families selected and analyzed as described in the text. The gene families are named as follows, vertically in each column. Column 1: Homeobox, Acetylcholine receptor, Actin, Apolipoprotein, Brain protein, Collagen, Colony stimulating factor; Column 2: Distal-less, Epidermal growth factor receptor, Even-skipped, Glucose transport, Interleukin, Interleukin receptor, Insulin; Column 3: a-Integrin, b-Integrin, Kirsten rat sarcoma, Keratin, Myosin heavy chain, Paired box; Column 4: Retinoic acid receptors, Rhombotin, Ribophorin, Retinoic acid X receptor, Sodium channel, T-cell receptor, Topoisomerase, Waved, Wingless.

Table 1. Possible orthologous genes linked to the *HOM* or Hox complexes in *C. elegans*, mouse, and human

C. elegans (Chromosome III)	Chromosome	
	Mouse	Human
ceh11 (homeobox group 9)	2,6,11,15	2,7,12,17
mab 5 (homeobox group 6)	6,11,15	7,12,17
Distal-less	2	2,7
Collagen	6, 11, 15	2,7,12,17
Ras family	2, 6, 11	2,7,12
Integrin, α subunit	2, 11, 15	2,12,17
Myosin heavy chain 3	11,15	7,17
Acetylcholine receptor	2,11	2,17
Protein kinase,ser/threo	11	17
Erbb	11	7,12,17
α-β-crystallin	11	2,17
Calcium channel, α-2b subunit	?	2,17
Synaptobrevin	6,11	12,17
GTP binding protein	6,11	7,12
ATP synthase, subunit (two)	?	2,12,17
Topoisomerase 2	2,11	17
Acyl CoA dehydrogenase	?	2,12
Engrailed	–	2,7
Vitronectin receptor, α subunit	2	2
Proton pump (Vpp-1)	11	2,17
Cyclic GMP phosphodiesterase	11	17
β-Spectrin	11	2
α-Spectrin	2	–
Calcium binding protein	11	–
Glutamate receptor	11	–
Notch repeat	2	–
Gelsolin	2	–

Data used in preparing this table were derived from the GDB, Human Genome Data Base accessed at Johns Hopkins University, and The Encyclopedia of the Mouse Genome accessed at Jackson Laboratory. *C. elegans* data were taken from Wilson et al. (1994). No attempt has been made to compare the *C. elegans* sequences with the possible orthologs located on any of the mouse or human chromosomes. The *HOM* complex in *C. elegans* is located on chromosome III. All *C. elegans* genes listed above are located within 2.2 megabases of the *HOM* complex. HOX clusters (*A, B, C, D*) are located on human chromosomes 7,17, 12, and 2, respectively. The locations of the mouse Hox clusters (*a, b, c, d*) are chromosomes 6, 11, 15, and 2, respectively. More than one mouse or human chromosome listed indicates the location of possible paralogs of these genes. – indicates that the gene is not located on a Hox cluster-bearing chromososome. ? indicates map location is not known.

(Forrester et al., 1987; Qian et al., 1991). A second possibility is that products of genes in the Hox cluster domains interact functionally, and thus are co-adaptive and require coordinated evolutionary modification in the sense of positive epistatic interactions. One can postulate that this can be most efficiently accomplished if the genes are in linkage and tend to assort together in populations. At present these and other notions must be regarded as highly speculative, but can serve as the bases of hypotheses to be tested by experimentation.

Hox CLUSTERS IN ECHINODERMS AND LOWER CHORDATES

The structure of the four Hox gene clusters in vertebrates implies that some Hox genes were lost following duplication (Kappen and Ruddle, 1993; Ruddle et al., 1994). Several possibilities exist concerning the history of cluster duplication (Kappen and Ruddle, 1993), and one of the simplest models involves the two-fold duplication of a single Hox gene cluster

are also located in proximity to the Hox gene clusters in *Caenorhabditis* (Table 1) and *Drosophila* (Table 2).

Assuming that the linkage pattern described above is indeed adaptive, one might speculate on its biological basis. One possibility is that the expression of the linked genes is regulated by *cis* regulatory effects that extend over large distances throughout the region bounding the Hox gene cluster. This is consistent with findings that enhancers can modulate gene activity over distances in the range of a hundred kilobases

Table 2. Possible orthologous genes linked to the *HOM* and *Hox* complexes in *D. melanogaster*, *C. elegans*, human, and mouse.

Drosophila	*C. elegans*	Mouse	Human
lab (homeobox group 1)	*C. elegans*	2,6,11	2,7,17
pb (homeobox group 2)		6,11	7,17
Dfd (homeobox group 4)	*C. elegans*	2,6,11,15	2,7,12,17
Scr (homeobox group 5)		6,11,15	7,12,17
Antp (homeobox group 6)	*C. elegans*	6,11,15	7,12,17
Ubx (homeobox group 7)		6,11	7,17
abd-A (homeobox group 8)		2,11,15	2,12,17
Abd-B (homeobox group 9-13)	*C. elegans*	2,6,11,15	2,7,12,17
Protein kinase, ser/thr	*C. elegans*	11	17
Actin		2,6	2,7,17
Tubulin	*C. elegans*		2,12
paired box-mesoderm		2,6	7
Ras proto-oncogene	*C. elegans*	2,6,11	2,7,12
Delta (Egf receptor-like)	*C. elegans*	11	7,12,17
Na$^+$, K$^+$ ATPase, α subunit		11	17
c-abl oncogene		2	–
Rhodopsin 2, 3, 4		6	–
Catalase		2	–
Arylsulfatase		15	–
U3, small nuclear RNA		11	–
U1, small nuclear RNA		11	–
Protein kinase, cAMP dep.	*C. elegans*	?	7,17
Casein kinase II, a subunit	*C. elegans*	?	–
Ribosomal protein L21	*C. elegans*	?	–
Calmodulin	*C. elegans*	–	7
RNA polymerase II, subunits		11	17
elongation factor 2a	*C. elegans*	?	?
empty spiracles		2	?
Nk homeobox		2	?

Data used in preparing this table were compiled from *Flybase* (on-line *Drosophila* database accessed by Gopher) and Lindsley and Zimm (1992). All of the *Drosophila* genes are located on chromosome 3 from 3-43 to 3-72. The split HOM-C complex is located at 3-47 and 3-59. Sources for mouse, human, and *C. elegans* data are listed in the legend of Table 1. *C. elegans* genes are all on chromosome III, and only include those genes located within 2.2 megabases of the HOM complex (Wilson, et al., 1994). Locations listed for human, mouse, and *C. elegans* indicate possible orthologs of the genes given for *Drosophila*, but not necessarily the exact ortholog of the *Drosophila* gene. – indicates that the gene is not located on a Hox cluster-bearing chromososome. ? indicates map location is not known.

containing ancestral representatives of all 13 paralogy groups, followed by relatively rapid loss of individual genes or cluster segments. Sequence conservation within the homeodomain allows us to estimate the number of Hox cluster genes (paralogy groups 1-10) in a species by the polymerase chain reaction (PCR; Frohman et al., 1990; Murtha et al., 1991; Pendleton et al., 1993). In cases where the number of Hox cluster genes is known, such PCR surveys have shown a recovery rate of more than 85% in a single sampling (Pendleton et al., 1993; and unpublished data).

Surveys of Hox cluster gene number in primitive chordates and echinoderms have provided some curious insights concerning the relationship between vertebrates and other deuterostomes. The hemichordate *Saccoglossus kowaleskii* most likely contains a single Hox gene cluster, with representatives from each of paralogy groups 1-9 (Pendleton et al., 1993). The Hox cluster number in the cephalochordate *Branchiostoma floridae* is more difficult to assess. Data from a PCR survey revealed the presence of 11 Hox cluster genes in amphioxus (Pendleton et al., 1993). On the basis of amino acid sequence similarity, the data predicted that these Hox genes

would be distributed over two clusters. Due to the short sequence (82 bp) amplified by PCR and the high similarity between some paralogy groups, paralogy assignments based on homology are necessarily speculative. Detailed linkage analysis of *B. floridae* Hox genes (Holland et al., this volume; Garcia-Fernàndez and Holland, personal communication) reveals a single Hox cluster containing representatives from each of paralogy groups 1-10. Comparison of the sequences from both data sets would be quite informative. Our data could be consistent with a single Hox cluster in *B. floridae*, although an argument could also be made for the existence of two genes in each of paralogy groups 1 and 3. Nineteen different Hox cluster genes were sampled in the agnathan *Petromyzon marinus*, suggesting that the closest extant relatives of the true vertebrates, the jawless fish, have at least two and most likely three or four Hox clusters (Pendleton et al., 1993).

Considering the important regulatory role that Hox cluster genes play during animal development (Shashikant et al., 1991; Ruddle et al., 1994), it is appropriate to examine Hox cluster structure in phyla that exhibit unique developmental qualities. The echinoderms comprise a deuterostome phylum that shows early developmental affinities with the the hemichordates, a phylum suspected to have close affinities with the chordates, but they have a radically different adult body plan that includes secondarily derived radial symmetry. Four Hox cluster genes have been previously isolated from the Hawaiian sea urchin *Tripneustes gratilla* using hybridization techniques (Dolecki et al., 1986, 1989; Wang et al., 1990). Three of these are related to 'medial' paralogy groups (Table 3), and one is most likely a member of 'posterior' paralogy group 9. Our own PCR survey of the sea urchin species *Strongylocentrotus purpuratus* and *Lytechinus variegatus* identified homologs of these *T. gratilla* genes, as well as three other Hox cluster genes (Table 3). One of these, *Hbox9*, is most likely a fourth homolog of 'medial' group genes. Two others, *Hbox7* and *Hbox10*, are most highly homologous to 'posterior' group 9, and do not appear to be related to 'posterior' paralogy groups 10-13. Interestingly, *Hbox7* and *Hbox10* are more closely related to one another than either are to posterior paralogy group 9. Paralogy groups 10-13 in amniotes are thought to have arisen by serial tandem duplication events beginning from paralogy group 9 (Kappen et al., 1993; Schubert et al., 1993). This mechanism might also explain the origin of *Hbox7* and *Hbox10* in echinoderms, but the lack of homology between these sequences and paralogy groups 10-13 in amniotes suggests that such duplications took place independently after the divergence of echinoderms and other deuterostomes.

Another unique distinction of sea urchin Hox clusters appears to be the curious absence of genes from 'anterior' paralogy groups 1-3. 'Anterior' Hox cluster genes have been reported for all other metazoan species examined (Except sponges, where no Hox cluster genes have yet been identified; Ruddle et al., 1994). One possible reason for failing to detect Hox cluster sequences in genomic DNA by PCR could be the presence of introns in the homeodomain. Introns that disrupt the homeodomain are rare, but the *Drosophila* 'anterior' Hox cluster genes *labial* and *proboscipedia* do contain homeodomain introns. None of the known vertebrate Hox cluster genes have introns in the homeodomain. The absence of 'anterior' Hox cluster genes in the echinoderms is also supported by the isolation of only four non-'anterior' genes by

Table 3. Sea urchin Hox cluster sequences

Gene name	Species	Derived sequence (aa 21-47)	Paralogy assignment
	T S L		
Hbox1	x x	HFNRYLTRRRRIELSHLLGLTERQIKI	'medial' groups 4-8
Hbox3	x x x	HFSRYVTRRRRFEIAQSLGLSERQIKI	'medial' groups 4-8
Hbox4	x x x	LFNMYLTRDRRLEIARLLSLTERQVKI	'posterior' group 9
Hbox6	x x	HYNRYLTRKRRIEIAQAVCLSERQIKI	'medial' groups 4-8
Hbox7	x x	QANMYLTRDRRSKLSQALDLTERQVKI	derived from 'posterior' group 9
Hbox9	x x	HFNRYLTRRRRIEIAHALGLTERQIKI	'medial' groups 4-8
Hbox10	x x	LYNMYLTRDRRSHISRALSLTERQVKI	derived from 'posterior' group 9

PCR surveys for Hox cluster sequences were carried out as described (Murtha et al., 1991; Pendleton et al., 1993) using genomic DNA from *S. purpuratus* (=S) and *L. variegatus* (=L). The resulting sequences were conceptually transcribed, and are listed in comparison to previously reported sequences from *T. gratilla* (=T) Hox cluster genes (Dolecki et al., 1986, 1989; Wang et al., 1990). The nomenclature used here is based on that used for *T. gratilla*. An 'x' in the species columns indicates those Hox cluster genes found for each species. Paralogy assignments are based on the best homologies within that sequence compared with other known Hox cluster genes from paralogy groups 1-13 (see text).

Table 4. Ascidian Hox cluster sequences

Gene name	Species	Derived sequence (aa 21-47)	Paralogy assignment
CI-1	*Ciona intestinalis*	HYNRYLTRRRRIEVAHTLCLTERQIKI	'medial' groups 4-8
SC-6	*Styela clava*	HFNRYLTRRRRIEIAHSLCLSERQIKI	'medial' groups 4-8
MO-4	*Molgula oculata*	HFNQYLTRERRLEVAKSVNLSDRQVKI	'posterior' group 10

PCR surveys for Hox cluster sequences were carried out as described by Murtha et al. (1991) and Pendleton et al. (1993), using genomic DNA. The listed sequences are conceptual translations of the PCR products. Paralogy assignments are based on the best homologies within that sequence compared with other known Hox cluster genes from paralogy groups 1-13 (see text).

hybridization in *T. gratilla* (see above). Also compelling is the failure to identify 'anterior' Hox cluster genes in sea urchin RNA (P. Martinez, personnal communication and unpublished data).

The urochordates are a large and varied subphylum, but share distinct developmental affinities with the chordates. The ascidians, for example, have a tadpole larval stage where they quite literally resemble what could be described as primitive chordates. The larvae are free-swimming and contain structures such as a notochord, dorsal nervous system, a primitive brain with paired sensory organs, mesenchyme cells, etc. (Jeffery and Swalla, 1992). After the larvae attach to a substrate, metamorphosis occurs generating a sessile, filter-feeding adult form rather unique among coelomates. A previously reported search for Hox cluster genes by hybridization in the ascidian *Halocynthia roretzi* (Saiga et al., 1991) failed to isolate any Hox cluster genes, as only one very diverged non-Hox cluster gene, *AHox1*, was described.

We have surveyed four different ascidian species for Hox cluster sequences by PCR. In each case, only one Hox cluster gene sequence could be identified (Table 4). Even more intriguing is the fact that the single Hox gene sequence detected in one ascidian species is entirely different from the ones present in other species. *Ciona intestinalis* surveys show one Hox cluster sequence related to 'medial' (Hox paralogs 4-8) class paralogy groups (Table 4), while *Styela clava* shows a clearly different 'medial' class Hox cluster sequence. Survey data from the more distantly related genus *Molgula* relate an even more curious tale. *Molgula oculata* represents a urodele ascidian species with a tailed larva from which a single Hox cluster gene related to 'posterior' paralogy group 10 was identified. The closely related species *Molgula occulta* represents

an anural ascidian which displays a tailless larval form. No Hox cluster sequences were detected in *M. occulta*, while the same non-Hox cluster genes were found that were present in the survey of *M. oculata* (unpublished data). Hybrids can be formed between these two *Molgula* species resulting in a hatched larva with a short tail (Swalla and Jeffery, 1990). In a review article on ascidian development, Jeffery and Swalla (1992) comment that anural development has probably evolved more than once in ascidians, suggesting that it may be the consequence of a relatively small number of loss-of-function mutations.

The urochordate Hox cluster data is indeed puzzling. A PCR survey of the pelagic (non-sessile) tunicate *Oikopleura dioica* (Holland et al., this volume) again is compatible with the presence of a single Hox cluster sequence, and in this case it is most closely related to 'anterior' paralogy group 1. It therefore appears that Hox cluster genes from 'anterior', 'medial' and 'posterior' paralogy groups are represented in this phylum. Have the bulk of urochordate Hox cluster gene sequences diverged beyond detection? Have wholesale changes in genomic organization (e.g. intron insertions) occurred? These scenarios seem unlikely since a different Hox cluster gene is present for each genus examined. One possible model considers that the forerunner of the urochordates had a single, complete Hox gene cluster. By any of a number of mechanisms of adaptation, the requirement for Hox cluster function in this species was lost. The single Hox cluster genes that remain in the different urochordate species may have been co-opted for different roles in the newly evolved developmental mechanisms. An example of the kind of role taken by these single Hox cluster genes may have already been alluded to, that is, could the single *Molgula* Hox cluster gene described above

be necessary for formation of the larval tail, with its loss resulting in anural development? Do any extant species exist that contain more than one Hox cluster gene? More rigorous examination of Hox cluster structure in the urochordates will be necessary to provide answers to these speculations.

DISCUSSION

In surveys for Hox cluster genes, representatives have been recovered for all of the major phyla with the sole exception of sponges. Moreover, a general correlation can be made between Hox gene number and the more recently evolved phyla (Ruddle et al., 1994). This result is consistent with the postulated conservative notion of development espoused by Ohno (1970). We refer to this process as the 'gene freeze' hypothesis. This hypothesis states that evolutionary innovations are facilitated by gene duplication of developmentally significant genes which allows the retention of old developmental functions and the introduction of new. The duplicated gene(s) then becomes integrated into the developmental plan of the organism and likewise becomes constrained with respect to change and is 'frozen'. This view of the developmental-evolutionary process has additional interesting properties. If we assume that developmental control is combinatorial and that developmental genes (eg., Hox genes) are interactive in the form of gene networks then the addition of genes by duplication may increase developmental possibilities geometrically. In other words, the introduction of a single gene may introduce a broad range of developmental possibilities and corresponding evolutionary options. Increasingly, evidence supports the view that evolution progresses in spurts (Gould and Eldredge, 1993). We submit that a discontinuous rate of evolutionary change is consistent with the gene freeze model. In this respect, the duplication of clusters of Hox genes in the antecedents of the vertebrates may have had important immediate consequences with respect to the vertebrate radiation.

Gene loss may also play an important role in the evolution of developmental mechanisms. In vertebrates, duplication events that created the four Hox gene clusters were presumably quickly followed by gene loss until extant cluster structures became 'frozen' (Kappen and Ruddle, 1993). As yet no remnants of the lost Hox cluster genes (e.g. pseudogenes) have been detected. It will be of great interest to examine the vertebrate classes in a detailed fashion with respect to the presence and absence of Hox cluster genes, since such data may possibly reveal the detailed patterns of gene loss, which in turn can give insight into class affinities. PCR surveys have provided compelling evidence that Hox cluster gene loss may also have occurred in other deuterostomes, particularly in the echinoderms and urochordates. Sea urchins appear to lack 'anterior' Hox cluster genes when they would be predicted to contain them on the basis of their phylogenetic position; they share a common ancestor with arthropods and chordates in which 'anterior' Hox cluster genes have been identified. It will be of interest to determine the extent to which 'anterior' Hox cluster gene loss, if true, has contributed to the unique anatomical characteristics shared by echinoderms. The ascidian Hox cluster gene data inspire curious speculation considering the central position accorded them with respect to vertebrate origins (Berrill, 1955). One historically prominent theory

posits that the ascidian larva, by means of a neotenic process, represents the ancestral form for vertebrate evolution (Garstang, 1894). The apparent loss of most Hox cluster genes and associated functions, if true, suggests that the urochordates are a derived group, diverging rather early from the stem lineage leading to the vertebrates. In addition, Hox cluster gene loss in urochordates, as speculated for the echinoderms, may have had profound consequences regarding developmental pathways and adaptive adult morphology.

The Hox clusters provide an exceptional system for the study of developmental processes, especially with regard to morphology. We have demonstrated the great potential of various deuterostome phyla as model systems to study the role of Hox cluster genes in the evolution of morphology, and the advancement to more complex forms. The detailed study of Hox cluster structure and its evolution will provide new and exciting insights into the origins of the vertebrates and the mechanism of their development.

REFERENCES

Beeman, R. W. (1987). A homoeotic gene cluster in the red flour beetle. *Nature (London)* **327**, 247-249.

Bentley, K. L, Bradshaw, M. S. and Ruddle, F. H. (1993). Physical linkage of the murine *Hoxb* cluster and nerve growth factor receptor on yeast artificial chromosomes. *Genomics* **18**, 43-53.

Berrill, N. J. (1955). *The Origin of the Vertebrates*. Oxford: Clarendon Press.

Buchberg, A. M. and Camper, S. A. (1993). Mouse Chromosome 11. *Mammalian Genome* **4 (Special Issue)**, S164-S175.

Bürglin, T. R. and Ruvkun, G. (1993). The *Caenorhabditis elegans* homeobox gene cluster. *Curr. Opin. Genet. Develop.* **3**, 615-620.

Bürglin, T. R., Ruvkun, G., Coulson, A., Hawkins, N. C., McGhee, J. D., Schaller, D., Wittmann, C., Müller, F. and Waterston, R. H. (1991). Nematode homeobox cluster. *Nature* **351**, 703.

Craig, I. and Craig, S. (1991). Genomic distribution of coding sequences homologous to those on chromosome 12. *Cytogenet. Cell Genet.* **58**, 1851-1852.

Craig, I. and Craig, S. (1992). Genomic distribution of coding sequences homologous to those on chromosome 12 paralogy map II. *Cytogenet. Cell Genet.* **61**, 255.

Dolecki, G. J., Wang, G. and Humphreys, T. (1989). Stage- and tissue-specific expression of two homeo box genes in sea urchin embryos and adults. *Nucl. Acids. Res.* **16**, 11543-11558.

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

Elandt-Johnson, R. C. (1971). *Probability Models and Statistical Methods in Genetics.* pp. 458-496. New York: John Wiley and Sons.

Ferguson-Smith, A. C. and Ruddle, F. H. (1988). The genomics of human homeobox-containing loci. *Pathol. Immunopathol. Res.* **7**, 119-126.

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

Forrester, W. C., Takegawa, S., Papayannopoulou, T., Stamatoyannopoulos, G. and Groudine, M. (1987). Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids. *Nucl. Acids Res.* **15**, 10159-10177.

Frohman, M. A., Boyle, M. and Martin, G. R. (1990). Isolation of the mouse *Hox-2.9* gene; analysis of embryonic expression suggests that positional

information along the anterior-posterior axis is specified by mesoderm. *Development* **110**, 589-607.

Garstang, W. (1894). Preliminary note on a new theory of the phylogeny of the chordata. *Zool. Anz.* **17**, 119.

Gould, S. J. and Eldredge, N. (1993). Punctuated equilibrium comes of age. *Nature* **366**, 233-227.

Hart, C. P., Compton, J. G., Langely, S. H., Hunihan, L., LeClair, K. P., Zelent, A., Chambon, P., Roderick, T. H. and Ruddle, F. H. (1992). Genetic linkage analysis of the murine developmental mutant velvet coat (*Ve*) and the distal chromosome 15 developmental genes *Hox-3.1*, *Wnt-1*, and *Krt-2*. *J. Exp. Zool.* **263**, 83-95.

Holland, P. W. H., Garcia-Fernàndez, J., Holland, L. Z., Williams, N. A. and Holland, N. D. (1993). The molecular control of spatial patterning in amphioxus. *J. Mar. Biol. Assoc. U.K.* **74**, 49-60.

Holland, P. W. H., Holland, L. Z., Williams, N. A. and Holland, N. D. (1992). An amphioxus homeobox gene: sequence conservation, spatial expression during development and insights into vertebrate evolution. *Development* **116**, 653-661.

Holland, P. W. H., Garcia-Fernàndez, J., Williams, N. A. and Sidow, A. (1994). Gene duplications and the origins of vertebrate development. *Development* Supplement, 125-133.

Jeffery, W. R. and Swalla, B. J. (1992). Evolution of alternate modes of development in ascidians. *BioEssays* **14**, 219-226.

Kappen, C. and Ruddle, F. H. (1993). Evolution of a regulatory gene family: HOM/HOX genes. *Curr. Opin. Genet. Dev.* **3**, 931-938.

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

Kappen, C., Schughart, K. and Ruddle, F. H. (1993). Early evolutionary origin of major homeodomain sequence classes. *Genomics* **18**, 54-70.

Kaufman, T. C., Seeger, M. A. and Olsen, G. (1990). Molecular and genetic organization of the *Antennapedia* gene complex of *Drosophila melanogaster*. *Adv. Genet.* **27**, 309-362.

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

Lindsley, D. L. and Zimm, G. G. (1992). *The Genome of Drosophila melanogaster*. San Diego: Academic Press.

Lundin, L. (1993). Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and house mouse. *Genomics* **16**, 1-19.

Mock, B. A., Neumann, P. E., Eppig, J. T., Duncan, R. C. and Huppi, K. E. (1993). Mouse Chromosome 15. *Mammalian Genome* **4** (Special Issue), S211-S222.

Moore, K. J. and Elliott, R. W. (1993). Mouse Chromosome 6. *Mammalian Genome* **4** (Special Issue), S88-S109.

Murtha, M. T., Leckman, J. F. and Ruddle, F. H. (1991). Detection of homeobox genes in development and evolution. *Proc. Natl. Acad. Sci., USA* **88**, 10711-10715.

Nadeau, J. H. (1989). Maps of linkage and synteny homologies between mouse and man. *Trends in Genetics* **5**, 82-86.

Nadeau, J. H. (1991). Genome duplication and comparative gene mapping. In *Advanced Techniques in Chromosome Research* (ed. K. Adolph), pp. 269-296. New York: Dekker.

Ohno, S. (1970). *Evolution by Gene Duplication*. Heidelberg: Springer-Verlag.

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

Qian, S., Capovilla, M. and Pirrotta, V. (1991). The *bx* region enhancer, a distant *cis*-control element of the *Drosophila Ubx* gene and its regulation by *hunchback* and other segmentation genes. *EMBO J.* **10**, 1415-1425.

Rabin, M., Ferguson-Smith, A., Hart, C. P. and Ruddle, F. H. (1986). Cognate homeobox loci mapped on homologous human and mouse chromosomes. *Proc. Natl. Acad. Sci., USA* **83**, 9104-9108.

Randazzo, F. M., Seeger, M. A., Huss, C. A., Sweeney, M. A., Cecil, J. K. and Kaufman, T. C. (1993). Structural changes in the *Antennapedia* complex of *Drosophila pseudoobscura*. *Genetics* **134**, 319-330.

Ruddle, F. H. (1989). Genomics and evolution of murine homeobox genes. In *The Physiology of Human Growth* (ed. J. M. Tanner and M. A. Preece), pp 47-66. Cambridge: Cambridge University Press.

Ruddle, F. H., Bartels, J. L., Bentley, K. L., Kappen, C., Murtha, M. T. and Pendleton, J. W. (1994). Evolution of *Hox* genes. *Ann. Rev. Genet.* **28**, In press.

Saiga, H., Mizokami, A., Makabe, K. W., Satoh, N. and Mita, T. (1991). Molecular cloning and expression of a novel homeobox gene *AHox1* of the ascidian, *Halocynthia roretzi*. *Development* **111**, 821-828.

Schubert, F. R., Nieselt-Struwe, K. and Gruss, P. (1993). The *Antennapedia*-type homeobox genes have evolved from three precursors separated early in metazoan evolution. *Proc. Natl. Acad. Sci., USA* **90**, 143-147.

Schughart, K., Kappen, C. and Ruddle, F. H. (1989). Duplication of large genomic regions during the evolution of vertebrate homeobox genes. *Proc. Natl. Acad. Sci., USA* **86**, 7067-7071.

Schughart, K., Kappen, C., and Ruddle, F. H. (1988). Mammalian homeobox-containing genes: Genome organization, structure, expression, and evolution. *Br. J. Cancer* **58** (Suppl. IX) 9-13.

Seimiya, M., Watanabe, Y. and Kurosawa, Y. (1992). Identification of homeobox-containing genes in the most primitive metazoa, sponge. *Zool. Sci.* **9**, 1213 (Abstr.).

Shashikant, C. S., Utset, M. F., Violette, S. M., Wise, T. L., Einat, P., Einat, M., Pendleton, J. W., Schugart, K. and Ruddle, F. H. (1991). Homeobox genes in mouse development. *Crit. Rev. Eukary. Gene Expres.* **1**, pp. 207-245.

Silver, L. M. (1993). Master locus list. *Mammal. Genome* **4** (Special Issue), S2-S9.

Siracusa, L. D. and C. M. Abbot, C. M. (1993). Mouse Chromosome 2. *Mammalian Genome* **4** (Special Issue), S31-S46.

Swalla, B. J. and Jeffery, W. R. (1990). Interspecific hybridization between an anural and urodele ascidian: Differential expression of urodele features suggests multiple mechanisms control anural development. *Dev. Biol.* **142**, 319-334.

Wagner, A. (1994). Evolution of gene networks by gene duplications: A mathematical model and its implications on genome organization. *Proc. Natl. Acad. Sci., USA* **91**, 4387-4391.

Wang, G. V., Dolecki, G. J., Carlos, R. and Humphreys, T. (1990). Characterization and expression of two sea urchin homeobox sequences. *Dev. Genet.* **11**, 77-87.

Wilson, R., Ainscough, R., Anderson et al. (1994). 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. *Nature* **368**, 32-38.