Expansion of the Hox Gene Family and the Evolution of Chordates

JW Pendleton, BK Nagai, MT Murtha, and FH Ruddle

PNAS 1993;90;6300-6304
doi:10.1073/pnas.90.13.6300

This information is current as of January 2007.

	This article has been cited by other articles: www.pnas.org#otherarticles
E-mail Alerts	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.
Rights & Permissions	To reproduce this article in part (figures, tables) or in entirety, see: www.pnas.org/misc/rightperm.shtml
Reprints	To order reprints, see: www.pnas.org/misc/reprints.shtml

Notes:

Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 6300–6304, July 1993
Evolution

Expansion of the *Hox* gene family and the evolution of chordates

JOHN W. PENDLETON, BRIAN K. NAGAI, MICHAEL T. MURTHA, AND FRANK H. RUDDLE*

Department of Biology, Yale University, P.O. Box 6666, New Haven, CT 06511-8112

Contributed by Frank H. Ruddle, March 29, 1993

ABSTRACT **Homeobox genes encode DNA-binding transcription regulators that participate in the formation of embryonic pattern or contribute to cell-type specificity during metazoan development. Homeobox genes that regulate axial patterning and segmental identity (*Hox/HOM* genes) share a conserved clustered genomic organization. Mammals have four clusters that have likely arisen from the duplication of a single ancestral cluster. The number of *Hox*-type genes in other deuterostomes was estimated by using a polymerase chain reaction sampling method. Increased *Hox* gene complements are associated with the appearance of chordate and vertebrate characters. Our data suggest the presence of one *Hox* cluster in the acorn worm, a hemichordate; two *Hox* clusters in amphioxus, a cephalochordate; and three in the lamprey, a primitive vertebrate.**

The increasing morphological complexity of the extant deuterostomes suggests a concomitant evolution of the developmental mechanisms underlying the morphological evolution of these groups. Deuterostomes share common features of embryogenesis, including radial cleavage and enterocoelous development. They include echinoderms, hemichordates, cephalochordates, and vertebrates, which represent wide variations in neural, muscular, and skeletal complexities. The echinoderms and hemichordates are relatively primitive forms that share several larval characteristics (1, 2). Hemichordates also share characters with the chordates, including the presence of pharyngeal arches and a dorsal nervous system, but lack a notochord, brain, and segmented musculature. The cephalochordate amphioxus has a notochord, dorsal nervous system, and segmented trunk, but it lacks characters such as complex paired sensory organs and a true brain (3), which are diagnostic for the vertebrates. Lampreys and hagfish, the most primitive vertebrates, are the only extant members of the vertebrate class Agnatha. The agnathans are notable for the first appearance of several vertebrate characters, including skull structures, a well-developed brain and anterior nervous system, and paired sensory systems (4, 5). However, agnathans lack jaws and paired fins (limbs), characters shared by all other vertebrates.

A class of DNA-binding transcription factors that contain a homeodomain plays a crucial role in the development of the protostomes and deuterostomes (6). The homeobox genes were initially identified in *Drosophila* (7, 8). Homologous genes have since been found in all other metazoan phyla examined to date (9). The *Hox* gene family is postulated to have increased in gene number by two distinct steps, expansion by gene duplication within a single cluster (10, 11) and cluster duplication (12). There is evidence (for reviews, see refs. 6 and 13 and references therein) to indicate that the homeobox genes regulate one another, serving as a gene network regulating development and cytodifferentiation. If the *Hox* gene system functions combinatorially, then increases in gene and cluster number might be expected to have a highly disproportionate developmental capacity to encode the formation of complex body plans (14). A prime objective in our investigation is to determine whether a positive relationship exists between *Hox* gene number and developmental complexity. We have investigated Hox genes and cluster number in a series of deuterostomes that exhibit different degrees of developmental complexity. An assessment of genetic homologies between these distantly related groups is made possible by the unique genomic properties of the homeobox genes.

Individual *Hox* genes in mice and humans on different clusters can be subdivided into 13 cognate groups on the basis of sequence similarity and by their positions within the clusters (15). Only a subset of the 13 *Hox* groups occur within a particular cluster, but their order on and among clusters is preserved without exception. Cluster amplification most probably occurred by the duplication of whole clusters with the subsequent loss of individual cognate genes (12). This is supported by the fact that genes within cognate groups are more similar with respect to sequence than are genes within a cluster. *Hox* genes are all transcribed from the same DNA strand, defining 5′ and 3′ ends of clusters. The order of genes within clusters also corresponds to the anterior boundaries of *Hox* gene expression along the anterior–posterior (A–P) axis during development. Thus, group 1 genes located at the 3′ ends of the clusters express at an anterior limit, whereas group 13 genes express at a posterior limit. Intervening groups express serially along the A–P axis. The four-cluster mouse system resembles that of the single *Drosophila* cluster (HOM-C), suggesting an ancestral *Hox/HOM* gene cluster (16, 17). Eight of the mammalian gene groups are similar in sequence to the *Drosophila* genes within the insect HOM-C cluster, and the relative positions of these genes are colinear with their counterpart positions in the mouse clusters. Moreover, the *Drosophila* genes are also expressed along the A–P axis in a manner that is consistent with their positions within their gene cluster. These highly conserved properties of the homeobox gene system, including sequence, position within clusters, and function, facilitate the identification of gene homologs in phylogenetically diverse species.

We have used the polymerase chain reaction (PCR) and degenerate primers directed towards the most highly conserved homeobox motifs to examine changes in the *Hox* gene family as morphological complexity has increased during evolution of deuterostomes. Our PCR screen has been successfully tested in *Drosophila* and in the mouse, two organisms in which the number of *Hox* genes and their genomic structure are already known. Our results from a hemichordate, the acorn worm *Saccoglossus kowalevskii*, the cephalochordate *Branchiostoma floridae* (amphioxus), and the agnathan *Petromyzon marinus* (lamprey)† show that the number of *Hox* genes is correlated with the evolutionary appearance of chordate and vertebrate characters. Furthermore, our data are consistent with the view that there is one

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*To whom reprint requests should be addressed.

†The sequences discussed in this paper have been deposited in the GenBank data base (accession nos. L14866–L14921).

Hox cluster in the acorn worm, two in amphioxus, and three or possibly four in the lamprey.

MATERIALS AND METHODS

Animals. Acorn worms, *Saccoglossus kowalevskii*, were obtained from the Marine Biological Laboratories (Woods Hole, MA). Amphioxus, *Branchiostoma floridae*, were obtained from Gulf Specimen Company (Panacea, FL). Larval lampreys, *Petromyzon marinus*, were collected from streams in central Connecticut. *Drosophila melanogaster* DNA was provided by S. Artavanis-Tsakonis (Yale University).

DNA Isolation. Either one or several animals (see below) were first anesthetized in benzocaine, and whole animals or selected tissues were ground in liquid nitrogen in a mortar and pestle. The resulting tissue was incubated at 50°C in a cell-lysis buffer of 100 mM EDTA/0.5% SDS/10 mM Tris·HCl, pH 8/100 mM NaCl/100 ng of proteinase K per ml for 3–15 hr, followed by extraction with phenol, phenol/chloroform/isoamyl alcohol (25:24:1, vol/vol), and chloroform/isoamyl alcohol (24:1, vol/vol). The solution was dialyzed for 24 hr against several changes of 10 mM Tris·HCl, pH 8/1 mM EDTA (TE) at 4°C. The resulting nucleic acids were treated with RNase and used directly for PCR, or further purified by precipitation with 0.1 vol of 3 M sodium acetate, pH 5.2, and 2 vol of ethanol, followed by resuspension in TE.

cDNA Preparation. Lamprey kidney RNA was isolated by a standard guanidinium thiocyanate/cesium chloride procedure (ref. 18, p. 196). The cDNA reaction was carried out with avian myeloblastosis virus reverse transcriptase (Promega), using the supplier's protocol and buffers and random hexamer primers. The cDNA reaction product was diluted and used as a substrate for PCR.

PCR. One hundred to 1000 ng of genomic DNA or various amounts of the diluted cDNA reaction mixture were used as a substrate for the PCR. Primers used were

HoxE: 5′-AAAGGATCCTGCAGARYTIGARAARGARTT-3′
HoxF: 5′-ACAAGCTTGAATTCATICKICKRTTYTGRAACCA-3′

(R = A or G, Y = C or T, K = G or T, and I = inosine). These primers are directed against highly conserved regions of the homeobox and are modified versions of the primers used by Murtha *et al.* (9) and contain restriction sites for *Bam*HI and *Hin*dIII. *Taq* DNA polymerase from Cetus was used with the supplier's buffer and a Mg^{2+} concentration of 1.5 or 3.0 mM. Primer concentration was 1 μM. PCR was carried out for 30–40 cycles; cycling parameters were 94°C for 1 min, 37°C for 1 min, and 72°C for 1 min, followed by a single cycle of 72°C for 7 min in either a Perkin–Elmer/Cetus or a Hybaid thermal cycler.

Cloning and Sequencing of PCR Products. PCR products were extracted once with chloroform and then digested with *Bam*HI and *Hin*dIII. The resulting PCR products of approximately 140 bp were gel purified and cloned in a modified pGEM4 (Promega) vector. Alternatively, PCR products were cloned in a TA vector from Invitrogen (San Diego). Inserts were sequenced by using standard dideoxynucleotide sequencing protocols and kits supplied by either Pharmacia or United States Biochemical.

Library Screening. A cosmid genomic library from the lamprey was constructed in the pWE15 cosmid vector (Stratagene). The library was screened with probes derived from PCR products amplified with primers HoxE and HoxF from genomic lamprey DNA. The positive clones were rescreened once or twice, and purified cosmid DNA was prepared by a standard alkaline lysis protocol (ref. 18, pp. 368–369) and used as a substrate for PCR as described above. Previously (19), a lamprey genomic phage library in the vector λ2001 provided by R. Doolittle (University of California, San Diego) had been screened with a *Drosophila* Antennapedia (*Antp*) probe, and a 1.2-kb genomic clone containing a homeobox was sequenced.

DNA Sequence Analysis. The informative sequence generated by our PCR protocol is 82 bp in length and encompasses 27 amino acids. The sequence includes the middle region of the homeobox from amino acids 21–47. Amplified sequences were compared with known homeobox sequences at both the nucleotide and deduced amino acid levels by using the FASTA program (20) and a data bank in our laboratory of approximately 300 nucleotide and 450 amino acid homeobox sequences, and assignments were made to cognate groups with the assistance of this comparison. In most cases, as can be seen in Table 1, assignment to cognate groups, based on a comparison to mouse homeobox sequences, was straightforward; however, in some instances, our assignment must be regarded as tentative, since much informative sequence lies at the 5′ and 3′ ends of the homeobox, in regions outside of the sequence amplified by our primers. In particular, the sequences belonging to cognate groups 6 and 7 have been grouped together, since it is not possible to distinguish between them on the basis of the 27-amino acid sequences we recover in our PCR screen. It should be noted that all of the sequences clearly belong to the *Hox* class of homeobox genes. We report data for cognate groups 1–10 only, since our primers do not amplify homeobox sequences from groups 11–13. The sequences reported in Table 1 represent PCR products that differ from one another by two or more nucleotides. Clones that differ by only one nucleotide are assumed to have arisen by PCR error or genetic polymorphism, both of which are observed at low frequency under the PCR conditions outlined above (M.T.M., unpublished data). A distance matrix for all of our clones was produced with the PAUP program (21).

RESULTS

We have tested our PCR screen on both mouse and *Drosophila* DNA to obtain a measure of its ability to recover known homeobox sequences. In the mouse (9), 24 sequences from cognate groups 1–10 were recovered out of the 27 expected in a sampling of 100 sequences. In our *Drosophila* survey, four different homeobox sequences were found in a sample of 61 sequenced clones that are identical to known members of the HOM-C complex. The numbers of individual clones found were as follows: *abd-A*, 10 clones; *Ubx*, 8 clones; *Dfd*, 2 clones; and *Antp*, 1 clone. *Scr* was not found, but the nonhomeotic genes *zen-1* (26 clones) and *ftz* (14 clones) were recovered. *zen-1* has one mismatch with our primer set, and *ftz* has none. The other homeotic members of HOM-C, *Abd-B*, *pb*, and *lab* have introns (22–24) that preclude their identification by our PCR screen. In *Drosophila*, we were therefore able to detect six of the seven HOM-C homeobox sequences expected to be amplified by our primer set. Thus, in *Drosophila* and in the mouse, the PCR screen used in this study detected 86% of the expected targets in the HOM-C complex and 89% of the expected targets in the four *Hox* clusters in the mouse. The high frequencies of recovery in such disparate species as *Drosophila* and the mouse suggest that recovery will be representative for other organisms as well.

Deduced amino acid sequences of the *Hox* clones from the acorn worm, amphioxus, and the lamprey are shown in Table 1. The total number of clones sequenced was 64 for the acorn worm, 75 for amphioxus, and 97 for the lamprey; of these clones, 48, 58, and 94, respectively, were *Hox* clones. The sequences demonstrate a broad distribution among cognate groups 1–10, and they reveal differences among the species in the numbers of clones assigned to individual cognate

Table 1. Comparison and homology assignments of *Hox* PCR clones from lamprey, amphioxus, and acorn worm

mouse gene	**mouse** HFNRYLTRRRRIEIAHALCLTERQIKI	**lamprey** HFNRYLTRRRRIEIAHALCLTERQIKI	**amphioxus** HFNRYLTRRRRIEIAHALCLTERQIKI	**acorn worm** HFNRYLTRRRRIEIAHALCLTERQIKI	*Dros*
Hoxa-1	---K----A--V---AS-Q-N-T-V--	--S-----A--V-V-A--Q-H-A-V-- a 5	-Y-K----A--V---A--N-N-T-V-- a 8*	---K----A--V---AM-G-N-T-V-- a 6	
Hoxb-1	---K--S-A--V---AT-E-N-T-V--	---K----A--V---A--Q-N-T-V-- b 3*	---K----A--V---A--Q-NHT-V-- b 1†		*lab*
Hoxd-1	---K----A------NC-Q-NDT-V--	--SK---CA--V---A--Q-N-T---- c 3			
		---K----A--V---A--Q-N-T---- d 1			
Hoxa-2	---K--C-P--V---AL-D-----V-V	---K--C-P-GV---AL-D-----V-V e 1‡	-Y-K-VCKP--K---SY-D-N---V-- c 3*	-Y-K--C-P------SM-D-S---V-V b 1	
Hoxb-2	---K--C-P--V---AL-D-----V-V		---K-VCKP--K---AY-D-N---V-- d 1*		*pb*
			-Y-K-VCKP--K---AC-D-N---V-- e 1		
			-Y-K-VCKP--K---AY-D-N---V-- f 1†		
Hoxa-3	------M-P--V-M-NL-N--------		------C-P--V-M-AM-N-------- g 3*	R-----QKQ--T-LSSV-N-------V c 5	
Hoxb-3	------C-P--V-M-NL-N-S------		------C-P--V-M-NL-N-----N-- h 1†		-
Hoxd-3	------C-P--V-M-NL-N--------				
Hoxa-4	----------T---S---V--	------A----V----S---S---V-- f 10†		-Y-Q----K------QMVN-------- d 1	
Hoxb-4	-Y---------V--------S------	----------S---S------ g 8			*Dfd*
Hoxc-4	-Y----------S---S------	----------L----S-A-S---V-- h 2			
Hoxd-4	----------T---P------				
Hoxa-5	----------S------	----------V-V-N----S------ i 6		----------G-S------ e 7	
Hoxb-5	----------S------	----------V-N----S------ j 5§			*Scr*
Hoxc-5	----------NN---N------				
Hoxa-6	----------N----------	----------V----S---------- k 6	---K----------G-------- i 14*	----------T-G-------- f 3	
Hoxb-6	-Y------------------------	----------V----S---------- l 3			*Antp*
Hoxc-6	----------N----------	----------V----S---------- m 1			
		---------- n 7*			
Hoxa-7	------------------------	---------- o 1	---K---------- j 6*	----------G-------- g 12	*Ubx*
Hoxb-7	-Y----------T----------	----------G-------- p 1			
Hoxb-8	L--P----K----VS---G-----V--	L--P----K----VS---G-----V-- q 4*	---K----K-------L-G-------- k 5*	-Y----------LS-L-G-------- h 2	*abd-A*
Hoxc-8	L--P----K----VS---G-----V--	L--P----K----VS-V-G-S---V-- r 2	----------S-G-------- l 3*		
Hoxd-8	L--P----K----VS-T-A-----V--				
Hoxa-9	L--M----D--Y-V-RL-N-----V--	L--M----D--Y-V-RG-N-----V-- s 4		L--M----E--VD--RL-N-----V-- i 11	
Hoxb-9	L--M----D--H-V-RL-N-S---V--	L--M----D--Y-V-RV-S-----V-- t 4†*			*Abd-B*
Hoxc-9	L--M----D--Y-V-RV-N-----V--	L--M----D--Y-V-RV-T-----V-- u 1			
Hoxd-9	L--M----D--Y-V-RI-N-----V--	L--M----D--Y-V-RG-N-----V-- v 1‡			
Hoxa-10	L--M----E--L--SRSVH--D--V--	L--M----E--L--SRGVN--D--V-- w 8	LY-M----E--Y--SQHVN-----V-- m 11*		
Hoxc-10	L--M----E--L--SKTIN--D--V--	L-SM----E--L--S-L-S--D--V-- x 4			-
Hoxd-10	L--M----E--L--SKSVN--D--V--				

* found in cosmid library
† found in both cDNA and DNA
‡ found only in cDNA
§ found in phage library

* found in both DNA preps
† found in "single" prep

Note that all known *Hox* genes from the mouse are included for reference purposes, not just those found in the study of Murtha *et al.* (9). All sequences are relative to the *Drosophila* Antennapedia sequence found at the top of the Table; a hyphen indicates an identical amino acid. The mouse gene and cognate groups are indicated in the left column and the *Drosophila* homologs in the right column. The number of individual clones found and sequenced for each sequence is indicated to the right of each sequence, as are our lab designations for each sequence.

groups. The PCR survey also identified sequences from a few non-*Hox* homeoboxes (data not shown).

Our data provide an estimate of cluster number. In organisms with four clusters, a maximum of four unique gene sequences is expected in any cognate group. The real number in some instances will be less than four, because of gene loss following cluster duplication. This can be seen in the mouse data set, where the actual number of genes is known from gene cloning (15). The variable distribution of genes among cognate groups in the mouse illustrates the difficulty of predicting cluster number solely on the basis of data for one or a few cognate groups. If only three different gene sequences were recovered by our PCR screen in mouse groups 4 and 9, then three clusters would be assumed when four actually existed. False negatives are expected to occur at a frequency of 10–15% (M.T.M., unpublished data) and, for that reason, we recovered at least 65 sequences for each species in our data set. Conversely, PCR errors and polymorphism may introduce false positives into our data set. We have examined the data sets critically with respect to these potential errors.

The lamprey data set was assembled from both genomic DNA and cDNA preparations (see Table 1). Four different clones were recovered that were similar to the group 1 (*lab*) sequences. Clones 1b and 1d differed by only 2 nucleotides, and they probably represent polymorphic forms at a single locus. In support of this interpretation, the mouse group 1 genes all differ from one another by 17 or 18 nucleotides. In the lamprey, six different group 6/7 clones were recovered, but several differed by only 2 or 3 nucleotides. Again, a good case can be made for polymorphism, and we can conclude that only four genes are actually present. Group 9 contains four different sequences, two of which (9s and 9v) differ by 4 nucleotides, have identical amino acid sequences, and come from different individuals, suggesting the recovery of allelic variants. Other combinations within this group show differences ranging from 8 to 10 nucleotides.

In the amphioxus data set, group 2 contains four different clones but all combinations show small differences of only 3–6 nucleotides. In these experiments, one DNA preparation was made from multiple individuals, while a second preparation originated from a single specimen (Table 1). Both samples yielded clones 2c and 2d, which show the maximum nucleotide difference. We conclude that amphioxus has one or, at the most, two loci representative of group 2 genes. It should also be noted that the amphioxus group 1 sequences differ by 20 nucleotides, group 3 by 21 nucleotides, group 6/7 by 14 nucleotides, and group 8 by 13 nucleotides. Interestingly, amphioxus clone 3g is also part of a gene (*AmphiHox3*) that was recently cloned from *B. floridae* (25). The acorn worm data set contained one gene for each cognate group 1–9.

An estimate of the number of loci present in each gene group for the sampled species after elimination of suspected false-positive clones is given in Table 2. The lamprey data are consistent with a three-cluster interpretation, but a four-cluster model cannot be rigorously excluded. The existence of only a few false-negative data could obscure the presence of four clusters. However, it should be noted that the total number of clones recovered, 19, and their distribution among the 10 groups (Table 2) are highly consistent with a three-cluster model. The amphioxus data are in good agreement with a two-cluster model, and the acorn worm results are strongly suggestive of only one cluster.

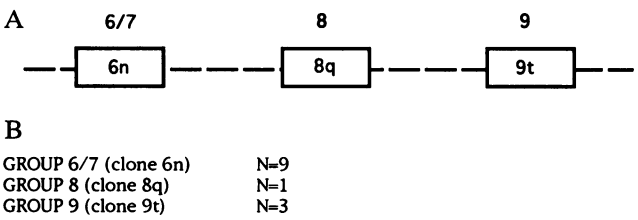

FIG. 1. Linkage of three *Hox* genes on cosmid *Pmcos2B*. PCR was done with HOX primers 3E and 5F on DNA from *Pmcos2B*. Thirteen minipreps were made from colonies identified as having inserts by PCR. (*A*) Putative linkage as revealed by PCR. The cognate groups (6/7, 8, and 9) are indicated above the laboratory names of the clones sequenced from *Pmcos2B*. All of these clones were found initially in the screen of lamprey genomic DNA. (*B*) Number of clones (*N*) found in PCR screen.

Analysis of a cosmid, *Pmcos2B*, isolated from a lamprey genomic library provides direct evidence that lamprey *Hox* genes are linked in a manner similar to that in mammals (Fig. 1). *Pmcos2B* was screened by PCR and was found to contain three homeoboxes within 30 kb, all of which had been previously identified in our genomic PCR screen (clones 6/7n, 8q, and 9t). By sequence comparison, these three clones had already been assigned to contiguous cognate groups (6/7, 8, and 9), supporting confidence in our overall assignments to cognate groups.

DISCUSSION

Our data reveal several important characteristics of *Hox* gene organization in animals that differ from one another in the number of chordate characteristics that they possess. First, our analysis demonstrates the presence of multiple *Hox* genes in the genomes of the acorn worm, amphioxus, and the lamprey. Second, we can assign these genes to the known cognate groups described in mice and humans on the basis of their amino acid and nucleotide signatures. These data provide indirect evidence that the *Hox* genes in these three animals are organized into clusters similar to those found in mammals. The cosmid data (Fig. 1) provide strong evidence that this is indeed the case in the lamprey.

Evidence for independent evolutionary pathways in cluster duplication is suggested by the present data. Amphioxus group 2 is interesting in this respect. We have tentatively interpreted the existence of four clones in this group as attributable to polymorphism. Another possibility is tandem gene duplication at this locus. It is also of interest that no genes were sampled in amphioxus that could be assigned to

Table 2. Estimated *Hox* cluster structure in lamprey, amphioxus, and acorn worm

Species	Cluster estimate	No. of genes in cognate group										Total
		1	2	3	4	5	6	7	8	9	10	
Mouse	4	3	2	3	4	3	3	2	3	4	3	30
Lamprey	3	2	1	0	3	2	4		2	3	2	19
Amphioxus	2	2	2	2	0	0	2		2	0	1	11
Acorn worm	1	1	1	1	1	1	2		1	1	0	9
Drosophila	1	*lab*	*pb*	—	*Dfd*	*Scr*	*Antp*	*Ubx*	*abd-A*	*Abd-B*	—	8

Our estimates of the number of genes in each cognate group for the lamprey, amphioxus, and acorn worm are listed for each group, as are our estimates of cluster number. The known *Hox* structure for the mouse and *Drosophila* are also indicated (see ref. 15 and references therein).

groups 4, 5, and 9. It is probable that some of these cognate groups have been lost altogether in amphioxus, since recovery of clones for those groups was uniformly good in all other species examined. In the acorn worm data set, it is noteworthy that all groups are represented except group 10. Groups 9 and 10 are especially interesting because the presence of group 10–13 genes in mammals has suggested serial tandem duplication events, beginning from group 9, that have extended the cluster serially in the 5′ direction. It is likely that this extension occurred prior to cluster duplication, since three of the four mammalian clusters show strong cognate group relationships involving groups 10–13. In our survey, group 10 genes were found in the lamprey and amphioxus, but not the acorn worm, suggesting that the serial extension of group 9 was underway in the common ancestor of the lamprey and amphioxus but not necessarily in that of the acorn worm.

An important finding of our study is the discovery of significant differences in the number and presumed structure of the *Hox* genes in the acorn worm, amphioxus, and the lamprey (Table 2). It is clear that there has been an increase in the number of these genes, which are known to act as important regulatory transcription factors during development. This increase is correlated with the evolutionary appearance of chordate and vertebrate characters. The PCR data also permit an estimate of cluster number in the sampled species. Analysis suggests that three clusters exist in the lamprey (although four cannot be rigorously excluded), two in amphioxus, and one in the acorn worm. These conclusions are borne out by a consideration of the number of *Hox* genes and, especially, by the distribution of clones among the individual cognate groups.

Studies on *Hox* gene mutations support the view that homeobox genes are involved in the ontogeny of structures that are important in vertebrate evolution. Loss-of-function mutations in the *Hoxa-1* gene in mice are associated with defects in structures derived from the neural crest, epidermal placodes, and branchial arches (26, 27). The *Hoxd-4* gene, when expressed ectopically in transgenic mice under the control of the *Hoxa-1* promoter, results in a homeotic posteriorization of the occipital bones of the skull (28). Interestingly, the authors describe the mutant phenotype as resembling, in part, the skull design of the adult lamprey. One of the hallmarks of chordate evolution is the emergence of the vertebrate head. Gans and Northcutt (29) have argued that the vertebrate head is a recent innovation, since elements relevant to its formation are absent from prechordate and primitive chordate forms. It is interesting to speculate that emergence of the complex head structures of the vertebrates has a basis in the amplification of the homeobox clusters. Comparative experimental studies of organisms with qualitative and quantitative differences in homeobox gene clusters may ultimately provide insight into this and related questions.

We thank Jean Wang for excellent technical help, and Leo Buss, John Reinitz, Mary Kritzer, C. S. Shashikant, Paz Einat, Janet Bartels, and Wendy Bailey for critically reading the manuscript. This work was supported by National Institutes of Health Grant GM009966 and National Science Foundation Grant IBN9119437.

1. Hyman, L. H. (1959) *The Invertebrates, Vol. V: Smaller Coelomate Groups* (McGraw–Hill, New York).
2. Barnes, R. D. (1980) *Invertebrate Zoology* (Saunders, Philadelphia).
3. Young, J. Z. (1981) *The Life of the Vertebrates* (Clarendon, Oxford, UK), 3rd Ed.
4. Hardisty, M. W. & Potter, I. C. (1971) *The Biology of Lampreys* (Academic, London), Vols. 1–5.
5. De Beer, G. R. (1937) *The Development of the Vertebrate Skull* (Clarendon, Oxford, UK).
6. Shashikant, C. S., Utset, M. F., Violette, S. M., Wise, T. L., Einat, P., Einat, M., Pendleton, J. W., Schugart, K. & Ruddle, F. H. (1991) *Crit. Rev. Eukaryotic Gene Exp.* **1**, 207–245.
7. McGinnis, W., Garber, R. L., Wirz, J., Kurouwa, A. & Gehring, W. J. (1984) *Cell* **37**, 403–408.
8. Scott, M. P. & Weiner, A. J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 4115–4119.
9. Murtha, M. T., Leckman, J. F. & Ruddle, F. H. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10711–10715.
10. Schughart, K., Kappen, C. & Ruddle, F. H. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7067–7071.
11. Schubert, F. R., Nieselt-Struwe, K. & Gruss, P. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 143–147.
12. Kappen, C., Schughart, K. & Ruddle, F. H. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5459–5463.
13. McGinnis, W. & Krumlauf, R. (1992) *Cell* **68**, 283–302.
14. Holland, P. (1992) *Bioessays* **14**, 267–273.
15. Scott, M. P. (1992) *Cell* **71**, 551–553.
16. Akam, M. (1989) *Cell* **57**, 347–349.
17. Graham, A., Papalopulu, N. & Krumlauf, R. (1989) *Cell* **57**, 367–378.
18. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
19. Pendleton, J. W., Chang, S., Cohen, M. J. & Ruddle, F. H. (1988) *Soc. Neurosci. Abstr.* **14**, 770.
20. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
21. Swofford, D. L. (1991) PAUP: *Phylogenetic Analysis Using Parsimony*, version 3.0k, computer program (Illinois Natural History Survey, Champaign, IL).
22. Mlodzik, M., Fjose, A. & Gehring, W. J. (1988) *EMBO J.* **7**, 2569–2578.
23. DeLorenzi, M., Ali, N., Saari, G., Henry, C., Wilcox, M. & Bienz, M. (1988) *EMBO J.* **7**, 3223–3231.
24. Kaufman, T. C., Seeger, M. A. & Olsen, G. (1990) *Adv. Genet.* **24**, 309–362.
25. Holland, P. W. H., Holland, L. Z., Williams, N. A. & Holland, N. D. (1992) *Development* **116**, 653–661.
26. Chisaka, O., Musci, T. S. & Capecchi, M. R. (1992) *Nature (London)* **355**, 516–520.
27. Lufkin, T., Dierich, A., LeMeur, M., Mark, M. & Chambon, P. (1991) *Cell* **66**, 1105–1119.
28. Lufkin, T., Mark, M., Hart, C. P., Dolle, P., LeMeur, M. & Chambon, P. (1992) *Nature (London)* **359**, 835–841.
29. Gans, C. & Northcutt, R. G. (1983) *Science* **220**, 268–274.