Int. J. Dev. Biol. 42: 617-620 (1998)

Short Contribution

Estimation of *Hox* gene cluster number in lampreys

ANNA C. SHARMAN[1] and PETER W.H. HOLLAND*

School of Animal and Microbial Sciences, The University of Reading, Reading, United Kingdom

ABSTRACT ***Hox*** **gene clusters are linked arrays of related homeobox genes with important roles in patterning the main body axis of animal embryos. Almost all invertebrates analyzed in detail, including a cephalochordate, have a single *Hox* gene cluster. In contrast, mammals have four such clusters inferred to have arisen by duplication. Data from other jawed vertebrates, including teleost fish, suggest they have at least four *Hox* gene clusters, implying that cluster duplication dates to very early in vertebrate evolution. Lampreys descended from one of the earliest vertebrate lineages and are thus critical in dating the duplication events. Here we analyze the *Hox* gene complement of a freshwater lamprey, *Lampetra,* using degenerate PCR. By analysis of the DNA sequences, deduced protein sequences, and by comparison to previous data from the distantly related sea lamprey, we conclude that lampreys have approximately 21 *Hox* genes from paralogous groups 1-10, plus a group 13 *Hox* gene. The data support the presence of three *Hox* gene clusters in lampreys more strongly than they support the presence of one, two or four gene clusters. We discuss how this situation may have arisen in evolution.**

KEY WORDS: *homeobox, hox cluster, evolution, gene cluster, vertebrate development*

Hox genes comprise a distinct class of homeobox genes found only in animal genomes. Their discovery, and the recognition of their conserved pattering roles in embryonic development, are perceived as being amongst the most important advances in modern biology (McGinnis *et al.*, 1984; reviewed by Gehring, 1993). One of the defining characteristics of true *Hox* genes is their arrangement into clusters of physically linked genes presumably resultant from tandem duplication of a primordial *Hox* gene early in animal evolution (Lewis, 1978; Kappen *et al.*, 1989; Scott, 1992). This clustered organization seems to be related to function, since the genes are generally expressed in a spatial and temporal order that is colinear with their physical order (Duboule and Dollé, 1989; Graham *et al.*, 1989). This ordered expressed plays a key role in the assignment of positional values to cells along the main body axis of an embryo, such that position can be translated into correct cell fate.

Mammals (including humans) have four clusters of *Hox* genes *(Hoxa* to *Hoxd)* on separate chromosomes, thought to have arisen by duplication from a single precursor cluster. Data from birds (Stein *et al.*, 1996), amphibians (Belleville *et al.*, 1992; Stein *et al.*, 1996) and teleost fish (Misof and Wagner, 1996; Aparicio *et al.*, 1997) suggest these vertebrates also have at least four *Hox* gene clusters, although additional clusters are present in some teleost fish (Prince *et al.*, 1998). All genes in the mammalian, bird, amphibian and teleost fish *Hox* gene clusters can be classified into thirteen paralogous groups (PG1 to PG13) by similarity of deduced protein sequence. Since no gene cluster has more than one member of each PG, these groups are thought to reflect descent by duplication from an ancestral cluster of 13 linked *Hox* genes (with one of each PG1 to PG13). It should be noted that the homeodomains encoded by homeoboxes of PG11 to PG13 are rather divergent in sequence from the more well known PG1-10 genes, particularly in the first alpha helix (Sharkey *et al.*, 1997), and have only been reported to date from jawed vertebrates.

In contrast to the multiple vertebrate clusters, all invertebrate species that have been surveyed in detail have just a single cluster of *Hox* genes (albeit secondarily split in two *Drosophila* species; Von Allmen *et al.*, 1996); the only putative exception to date being the horseshoe crab *Limulus* (Cartwright *et al.*, 1993). Interestingly, even the invertebrates most closely related to the vertebrates (cephalochordates or amphioxus) have just a single *Hox* gene cluster, with representatives of (at least) PG1 through to PG10 (Garcia-Fernàndez and Holland, 1994). Cephalochordate genes from PG11 to PG13 have not been reported.

Abbreviations used in this paper: PCR, polymerase chain reaction; PG, paralogous group.

***Address for reprints:** School of Animal and Microbial Sciences, The University of Reading, Whiteknights, Reading RG6 6AJ, UK. FAX: +44 (0)118 9316644. e-mail: p.w.h.holland@reading.ac.uk

[1]**Present address:** Institute of Neurobiology, University of Heidelberg, In Neuenheimer Feld 364, D-69120 Heidelberg, Germany.

0214-6282/98/$10.00

Printed in Spain

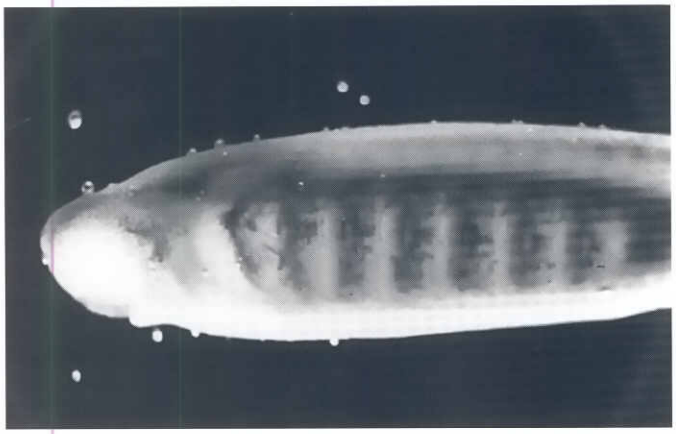

Fig. 1. Lamprey ammocoete. *Lampetra ammocoete from the River Pang, Berkshire.*

It has been suggested that expansion of the *Hox* gene family was a key event in vertebrate evolution, perhaps permitting the evolution of greater complexity in vertebrate embryonic development (Pendleton *et al.*, 1993; Garcia-Fernàndez and Holland, 1994). Deducing the precise timing and mechanism of duplication from one *Hox* gene cluster to four will help test this hypothesis and aid elucidation of the evolutionary consequences of *Hox* gene family expansion. This necessitates analysis of *Hox* genes in taxa that diverged from the vertebrate lineage after the divergence of cephalochordates but before the origin of bony fish.

We present here the cloning of *Hox* gene homeoboxes from the brook lamprey, *Lampetra planeri*, using degenerate PCR. This strategy usually suffers from the pitfall of incomplete sampling of genes, but here we overcome this problem by comparison to a distant lamprey species. The results are consistent with the presence of three clusters of *Hox* genes in lampreys.

PCR cloning of lamprey Hox genes

We used PCR amplification to survey the *Hox* gene compliment of the brook lamprey (Fig. 1). The first set of primers used, SO1* and SO2, are designed to amplify DNA encoding amino acids 21-47 from Hox homeoboxes of the divergent group PG13, thought to play a role in patterning the most posterior part of the body axis. Amplification with SO1* and SO2 generated 16 lamprey clones, of which five were homeoboxes. Four showed close sequence similarity to vertebrate PG13 genes, while one was closer to PG10. This is the first detection of a PG13 *Hox* gene outside the higher (jawed) vertebrates, and significantly pushes back the date for origin of this group of *Hox* genes. It implies that the *Hox* gene cluster expanded (by tandem gene duplication) to thirteen distinct genes even before the divergence of lamprey and jawed vertebrate lineages, approximately 450 million years ago.

The second set of primers used, SO1 and SO2, are designed to amplify the equivalent coding region, but from the more well known *Hox* gene groups PG1 to PG10. After amplification with these primers, we determined the DNA sequence of 51 recombinant clones; 35 contained a single homeobox and two contained two homeoboxes (total of 39 homeoboxes from SO1 and SO2, plus 5 from SO1* and SO2).

Sequences with up to two nucleotide differences over 82 sites were inferred to derive from the same gene (alleles or Taq polymerase errors); greater than two differences reflect different genes. This allowed the recognition of 18 distinct brook lamprey *Hox* genes.

The deduced amino acid sequences were then compared with those of jawed vertebrate homeobox genes, to classify the clones into paralogous groups. For most sequences, it was straightforward to assign clones to a particular PG; however, the high level of sequence similarity between homeoboxes of PG4 to PG7 precluded conclusive assignment. Clones from these PGs were therefore designated as "PG4-7". These results confirm that the brook lamprey has at least 17 *Hox* genes from PG1 to PG10, representing PG1, PG2, PG3, PG8, PG9, PG10 and the PG4-PG7 groups, plus the PG13 gene. A divergent clone with similarity to the Chox-7/Gbx family was also obtained (not shown). Deduced amino acid sequences, and allocation to paralogous groups, are shown in Figure 2; nucleotide sequences are available through the EMBL and GenBank databases (accession numbers AF044797 to AF044814).

Hox many Hox genes in lampreys?

PCR using degenerate primers is a powerful strategy to isolate genes from particular gene families. The general strategy may be criticized as not revealing the full diversity of a gene family, since amplification bias may lead to some genes being missed. In the present study, we were in the unusual position of being able to test

PG	Antp	HFNRYLTRRRRIEIAHALCLTERQIKI	Number clones
Hox-1	mouse consensus	...K..s.A..v...Ax.Q.n.T.V..	
	LpHox1A	..S.....A..V.V.A..Q.H.A.V..	2
	LpHox1B	...K....A..V...A..Q.N.T.V..	2
	LpHox1C	..SK....A..V...A..Q.N.T....	4
Hox-2	mouse consensus	...K..C.P..V...AL.D.....V.V	
	LpHox2A	...K..C.P..V...AL.D.....V.V	4
Hox-3	mouse consensus	......c.P..V.M.NL.N........	
	LpHox3A	......C.P..V.M.NL.N........	1
Hox-4	mouse consensus	.y..............t...S.....	
Hox-5	mouse consensus	....................s.....	
Hox-6	mouse consensus	..............n..........	
Hox-7	mouse consensus	.y..............t..........	
	LpHox4-7A	.............V.N....S.....	2
	LpHox4-7B	....	4
	LpHox4-7C	..........V....S..........	1
	LpHox4-7D	......A....V....S...S...V..	1
	LpHox4-7E	..........V.V.S....S.....	1
Hox-8	mouse consensus	L..P....K....VS...g.....V..	
	LpHox8A	L..P....K....VS.V.G.S...V..	2
	LpHox8B	L..P....K....VS...G.....V..	4
Hox-9	mouse consensus	L..M....D..y.V.Rl.N.....V..	
	LpHox9A	L..M....D..Y.V.RG.N.....V..	1
	LpHox9B	L..M....D..Y.V.RV.S.....V..	3
	LpHox9C	L..M....D..Y.V.RV.T.....V..	1
Hox-10	mouse consensus	L..M....E..L..Sksvn..D..V..	
	LpHox10A	L.SM....E..L..SRL.S..D..V..	2
	LpHox10B	L..M....E..L..SRGVN..D..V..	5
Hox-13	mouse consensus	Ax.KFI.KdK.Rr.SAxTn.S...vT.	
	LpHox13A	AA.KFI.KDK.RK.SA.IS.S...VT.	4

Fig. 2. *Lampetra* homeodomains. *Assignment of L. planeri Hox genes to paralogous groups (PG). Deduced amino acid sequences are compared to consensus sequences of the mouse Hox genes from PG1-10 and 13, shown relative to the Antp homeodomain of Drosophila. In consensus sequences, a capital letter indicates a residue that is the same in all mouse genes from that PG; a lowercase letter is represented in the majority of genes; x indicates a more variable site. Dots indicate residues identical to Antp. DNA sequences have EMBL/GenBank accession numbers AF044797 to AF044814.*

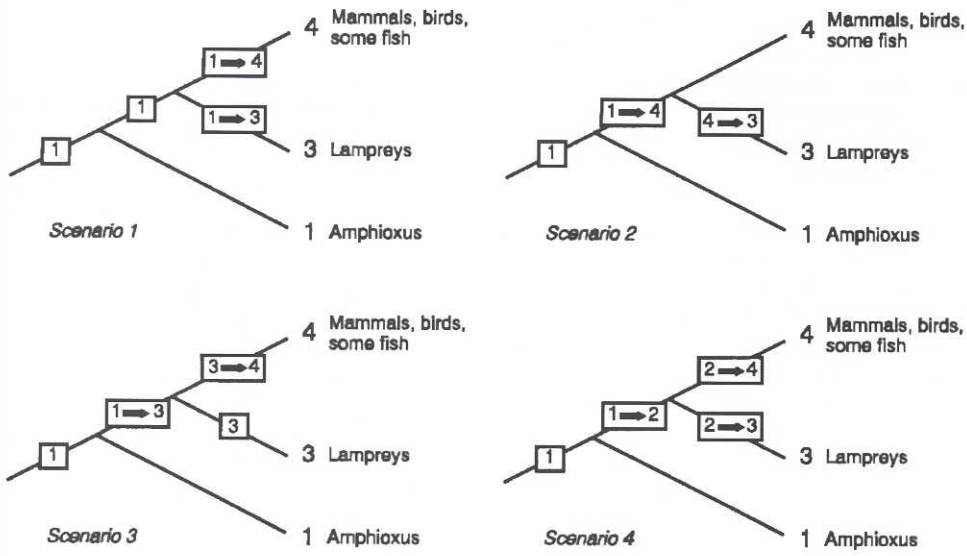

Fig. 3. *Hox* cluster evolution. *Four alternative evolutionary scenarios compatible with current estimates of Hox gene cluster number in amphioxus, lampreys and jawed vertebrates. Ancestral states, gene cluster duplications or gene cluster losses shown in boxes. See text for details.*

this criticism and assess how close to saturation is our survey of lamprey *Hox* genes. Our test is based on the principle that two species from the same higher level taxon (i.e., two lampreys) are expected to have largely the same *Hox* gene complement, but if amplification bias is significant it should affect different genes in different species due to divergence of DNA sequences. Pendleton *et al.* (1993) have previously reported the cloning of *Hox* gene sequences from the sea lamprey, *Petromyzon marinus*; taxonomically, this is a very distant lamprey species from the one used in this study, belonging to a separate genus and family. Pendleton *et al.* sequenced 94 *Hox* gene fragments, and assigned these to 20 genes. Our examination of the nucleotide sequences agrees with this conclusion. We also examined the deduced amino acid sequences of each *P. marinus* gene to assign each to a PG; we agree with all the lamprey PG assignments suggested by Pendleton *et al.* except that we argue PG4-7 cannot be distinguished and should be considered together. In conclusion, the *P. marinus* PCR survey yielded 3 genes from PG1, 1 from PG2, 9 from PG4-7, 2 from PG8, 3 from PG9 and 2 from PG10.

If PCR amplification bias has seriously distorted sampling of *Hox* genes in the two species, the 17 *Hox* genes from PG1-10 identified from *L. planeri* should have limited congruence with the 20 *Hox* genes from PG1-10 identified from *P. marinus*. Deduced amino acid sequences were therefore compared between the two species. We found remarkable congruence between the data sets. Of the 17 *L. planeri Hox* genes identified, 16 correspond to *P. marinus* genes with either one or no amino acid changes. Thus, only four *Hox* genes were found uniquely in the *P. marinus* survey, and only one PG1-10 gene was unique to the *L. planeri* survey (Table 1).

We conclude that the two independent PCR screens of different lamprey species both approached saturation, with little amplification bias. Lampreys are deduced to have approximately 21 (or very few more) *Hox* genes from PG1-PG10.

How many Hox gene clusters?

In all higher vertebrates examined to date, no examples have been found of two genes from the same paralogous group existing within a single *Hox* gene cluster. Each cluster contains one representative of each PG, or no representative where the gene is deduced to have been lost. Hence, the maximum number of genes assignable to a single PG is a powerful clue to the number of *Hox* gene clusters present.

Table 1 shows the number of *L. planeri* and/or *P. marinus Hox* genes assignable to each PG (except that PG4-7 are amalgamated for reasons discussed above). It can be seen the maximum number of genes in a PG is three. This suggests the presence of three *Hox* gene clusters in lampreys.

An alternative method to estimate the number of *Hox* gene clusters is to calculate the number of gene losses that must have occurred, following gene cluster duplication, to yield the spectrum of genes observed. For example, it is formally possible than an organism with four *Hox* gene clusters could possess just two or three *Hox* genes in every PG, due to a large number of gene losses after cluster duplication. The mouse and the pufferfish are the two vertebrates for which *Hox* gene clusters have been best characterized. The mouse has 39 *Hox* genes arranged in four clusters (Zelster *et al.*, 1996); 30 of these genes represent PG1 to PG10. Four *Hox* gene clusters have also been characterized in a pufferfish; here 24 genes represent PG1 to PG10 (Aparicio *et al.*, 1997). By definition, PG1 to PG10 must have contained 40 genes immediately after duplication to produce four *Hox* clusters. Hence, the rate of gene loss has been 25% in mouse and 40% in pufferfish. We argue that lampreys have 21 *Hox* genes from PG1 to PG10. We calculate, therefore, that if these represent four *Hox* gene clusters, the rate of gene loss would be 50%. Three clusters would equate to a 33% rate of gene loss, whilst two clusters would indicate a 0% rate (and be very hard to reconcile with PG membership, see above). This argument marginally favours the presence of three *Hox* gene clusters in lampreys, supporting the conclusion from PG membership above.

Evolutionary significance of lamprey Hox gene clusters

The multiple *Hox* gene clusters of mammals, birds, amphibia and teleost fish have clearly evolved from duplication of an ancestral *Hox* gene cluster. When did this occur? Since three of the pufferfish *Hox* gene clusters are clearly orthologous to three of the mammalian clusters *(Hoxa, Hoxb, Hoxc)*, the cluster duplication

events must predate the divergence of ray-finned fish and tetrapod lineages (approximately 408 million years ago). The finding of a single, complete, *Hox* gene cluster in a cephalochordate refines the timing of these events to after the divergence of cephalochordate and vertebrate lineages (over 520 million years ago). Hence, a single *Hox* gene cluster duplicated to give four clusters between approximately 520 and 408 million years ago.

What are the implications of detecting three *Hox* gene clusters in lampreys (a lineage that branched from the other vertebrates approximately 450 million years ago)? At least four alternative scenarios can be envisaged (Fig. 3). First, independent duplication. Under this scenario, the latest common ancestor of lampreys, ray-finned fish and tetrapods had just a single *Hox* gene cluster; this duplicated independently to give three clusters in lampreys and four in the ancestor of ray-finned fish and tetrapods. Note that this model implies that the origin of vertebrates occurred independently of *Hox* gene cluster duplication. Second, gene cluster loss. In this model, the duplication from a single *Hox* gene cluster to four occurred in the very earliest vertebrates, and the lampreys have simply lost an entire *Hox* gene cluster. Third, descent from an intermediate of three. This suggests that duplication from one to three *Hox* gene clusters predated the separation of the lamprey and jawed vertebrate lineages. The former lineage retained this intermediate state, whilst the latter lineage encountered an additional cluster duplication yielding four. Finally, descent from an intermediate of two. Under this scenario, the ancestral *Hox* gene cluster duplicated to give two gene clusters, before the separation of the lamprey and jawed vertebrate lineages. The former lineage then underwent a single additional duplication to give three, whilst in the latter lineage the two clusters doubled to give four.

At present, *Hox* gene sequences provide insufficient data to distinguish between these alternatives, although we note that each makes different predictions for the molecular phylogenetic relationships between the *Hox* gene clusters of cephalochordates, lampreys, teleost fish and mammals. However, we favor the last hypothesis (that the common ancestor of lampreys and jawed vertebrates possessed two *Hox* gene clusters) since evidence from other gene families argues that tetraploidy caused a doubling of all genes close to the origin of jawed vertebrates (Sharman and Holland, 1996).

Experimental Procedures

Lamprey ammocoetes of the genus *Lampetra* were collected from the River Pang, Berkshire, UK (Fig. 1). We consider these to be brook lampreys, *L. planeri*, although it should be noted that at the stage of development used *L. planeri* is not confidently distinguishable for *L. fluviatilis*. Total DNA was extracted by a standard proteinase K/SDS method (Holland, 1993). PCR amplification used either primer SO1 (GARCTNGARAARGARTT) or SO1* (GARCTNGARAARGARTA), with SO2 (CKNCKRTTYTGRAACCA). Cycling conditions and cloning strategies followed Holland (1993).

Acknowledgments

This work was funded by the BBSRC. We thank Mr J. Emm for permission to collect lampreys from his land.

References

APARICIO, S., HAWKER, K., COTTAGE, A., MIKAWA, Y., ZUO, L., VENKATESH, B., CHEN, E., KRUMLAUF, R. and BRENNER, S. (1997). Organization of the Fugu rubripes Hox clusters: evidence for continuing evolution of vertebrate Hox complexes. *Nature Genet. 16:* 79-83.

BELLEVILLE, S., BEAUCHEMIN, M., TREMBLAY, M., NOISEUX, N. and SAVARD, P. (1992). Homeobox-containing genes in the newt are organized in clusters similar to other vertebrates. *Gene 114:* 179-186.

CARTWRIGHT, P., DICK, M. and BUSS, L.W. (1993). HOM/Hox type homeoboxes in the chelicerate Limulus polyphemus. *Mol. Phylogenet. Evol. 2:* 185-192.

DUBOULE, D. and DOLLÉ, P. (1989). The structural and functional organization of the murine *Hox* gene family resembles that of *Drosophila*. *EMBO J. 8:* 1497-1505.

GARCIA-FERNÀNDEZ, J. and HOLLAND, P.W.H. (1994). Archetypal organization of the amphioxus Hox gene cluster. *Nature 120:* 407-413.

GEHRING, W.J. (1993) A history of the homeobox. In *Guidebook to the homeobox genes* (Ed. D. Duboule) Oxford University Press, Oxford, pp. 3-10.

GRAHAM, A., PAPALOPULU, N. and KRUMLAUF, R. (1989). The murine and Drosophila homeobox gene complexes have common features of organization and expression. *Cell 57:* 367-378.

HOLLAND, P.W.H. (1993). Cloning genes using the polymerase chain reaction. In *Essential Developmental Biology: A Practical Approach* (Ed. C.D. Stern and P.W.H. Holland). IRL Press, Oxford, pp. 243-255.

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.

LEWIS, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature 276:* 565-570.

MCGINNIS, W., LEVINE, M.S., HAFEN, E., KUROIWA, A. and GEHRING, W.J. (1984). A conserved DNA sequence in homeotic genes of *Drosophila* Antennapedia and Bithorax complexes. *Nature 308:* 428-433.

MISOF, B.Y. and WAGNER, G.P. (1996). Evidence for four Hox clusters in the killifish Fundulus heteroclitus (Teleostei). *Mol. Phylogenet. Evol. 5:* 309-322.

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.

PRINCE, V.E., JOLY, L., EKKER, M. and HO, R.K. (1998). Zebrafish hox genes: genomic organization and modified colinear expression patterns in the trunk. *Development 125:* 407-420

SCOTT, M.P. (1992). Vertebrate homeobox gene nomenclature. *Cell 71:* 551-553.

SHARKEY, M., GRABA, Y. and SCOTT, M.P. (1997). Hox genes in evolution: protein surfaces and paralog groups. *Trends Genet. 13:* 145-151.

SHARMAN, A.C. and HOLLAND, P.W.H. (1996). Conservation, duplication and divergence of developmental genes during chordate evolution. *Neth. J. Zool. 46:* 47-67.

STEIN, S., NISS, K. and KESSEL, M. (1996). Checklist - vertebrate homeobox genes. *Mech. Dev. 55:* 519-533.

VON ALLMEN, G., HOGGA, I., SPIERER, A., KARCH, F., BENDER, W., GYURKOVICS, H. and LEWIS, E. (1996). Splits in fruitfly Hox gene complexes. *Nature 380:* 116.

ZELSTER, L., DESPLAN, C. and HEINTZ, N. (1996). Hoxb-13 - a new Hox gene in a distant region of the Hoxb cluster maintains colinearity. *Development 122:* 2475-2484.

Received: September 1997

Accepted for publication: November 1997