

# Hox Gene Complexity in Medaka Fish May Be Similar to That in Pufferfish Rather Than Zebrafish

Gene Kurosawa,\* Kouji Yamada,† Hiroshi Ishiguro,† and Hiroshi Hori\*<sup>1</sup>

\*Division of Biological Sciences, Graduate School of Science, Nagoya University, Nagoya 464-8602, Japan; and

†Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Aichi 470-1192, Japan

Received May 14, 1999

**Changes in number and the genomic organization of *Hox* genes have played an important role in metazoan body-plan evolution. They make cluster(s), and in vertebrates, each cluster contains different number of *Hox* genes that have been classified into 13 groups. There are 39 *Hox* genes in four clusters on different chromosomes in the mammalian genome. In the fish, while 31 *Hox* genes in four clusters have been identified in pufferfish *Fugu rubripes*, 47 *Hox* genes in seven clusters exist in the zebrafish *Danio rerio*. To estimate the evolutionary origin of *Hox* organization in ray-finned fishes, we searched for *Hox* genes in the medaka fish *Oryzias latipes*, with a taxon thought to be widely separated from those of pufferfish and zebrafish. We synthesized various mixed oligonucleotides that can work as group-specific primers for PCR, then cloned and sequenced amplified fragments. Numbers of *Hox* genes identified in the present study were 2 for group 1, 2 for group 2, 1 for group 3, 3 for group 4, 6 for groups 5–7, 2 for group 8, 4 for group 9, 3 for group 10, 1 for group 12, and 3 for group 13. The primers specific for group 11 did not function in this study. Thus, at least 27 *Hox* genes are present in medaka genome, suggesting that the *Hox* gene complexity of the medaka genome is similar to that of the pufferfish rather than the zebrafish.** © 1999 Academic Press

**Key Words:** *Hox* genes; medaka fish; *Oryzias latipes*.

Diversification with in case in complexity of organisms in evolution has occurred mainly by multiplication of preexistent genes and/or chromosomes (1). *Hox* genes, whose products are transcription factors that play major roles in specifying cell fate during developmental processes of animals, are typical examples (2). *Hox* genes, identified in all metazoans except for the

sponge (3), are encoded on chromosomes in clusters with the order along the chromosome reflecting the order they act along the body (4). Such clustering and ordering of *Hox* genes have been described for *Drosophila*, mammals, *C. elegans* and even cnidaria (5–7). All the *Hox* genes in one cluster are presumed to have been generated by gene duplication followed by diversification and natural selection (8). In *Drosophila*, eight *Hox* genes (termed *HOM-C* in *Drosophila*) named *lab*, *pb*, *Dfd*, *Scr*, *Antp*, *Ubx*, *abdA* and *AbdB*, have been identified (5). *Hox* genes in mammals are present in four separate clusters and 39 *Hox* genes have been isolated from human and the mouse (5). They are classified into 13 groups based on sequence similarities (9). Since there exists only one cluster composed of 10 *Hox* genes (group 1 to 10) in *Amphioxus* (10), multiplication of an ancestral cluster containing 13 *Hox* genes should have occurred in vertebrate evolution (11). In the pufferfish *Fugu rubripes*, 4 *Hox* clusters that encode 31 *Hox* genes in total have been identified (12). The number of clusters is the same as that in mammals although the number of *Hox* genes in each cluster is slightly different. Recently, however, Amores *et al.* (13) reported the presence of 7 *Hox* clusters containing 47 *Hox* genes in zebrafish *Danio rerio*. Thus, multiplication of *Hox* clusters might have occurred differently among fish species. It has been argued that multiplication of *Hox* clusters is not restricted to chromosome loci encoding *Hox* genes. Also large DNA regions containing *wnt*, *hedgehog*, *dlx* and *en* appear to have been multiplied together (14). Therefore, it is of interest to analyze *Hox* clusters in other fish species.

In the present study, to estimate the evolutionary origin of the *Hox* organization in ray-finned fishes (Subclass Actinopterygii), we searched for *Hox* genes in medaka fish *Oryzias latipes* (order Atheriniformes), with a taxon thought to be widely separated from those of pufferfish (order Tetraodontiformes), and zebrafish (order Cypriniformes). Results of PCR suggest a *Hox* gene complexity more similar to the pufferfish case.

<sup>1</sup> To whom correspondence should be addressed. Fax: +81-52-789-2974. E-mail: [hori@bio.nagoya-u.ac.jp](mailto:hori@bio.nagoya-u.ac.jp).

Abbreviations used: BAC, bacterial artificial chromosome; bp, base pair(s); kb, kilobase(s); MF, medaka fish; MM, mammals; PCR, polymerase chain reaction; PF, pufferfish; ZF, zebrafish.

5' primer			
Group-specific primer			
Group	Nucleotide sequences	Amino acid sequences	Positions
1	5'- <u>GGGGATCC</u> (A/T/G/C)T(A/T/G/C)(C/A)G(A/G)AC(A/T/G/C)AA(T/C)TT-3'	(I/V/L)RTNF	4-8
2-5	5'- <u>GGGGATCC</u> (A/C)G(A/G)AC(A/T/G/C)GC(A/T/G/C)TA(T/C)AC-3'	RTAYT	5-9
5-7	5'- <u>GGGGATCC</u> TA(T/C)AC(A/T/G/C)(A/C)G(A/G)TA(T/C)CA(A/G)AC-3'	YTRYQT	8-13
6(C)	5'- <u>GGGGATCC</u> (A/C)G(A/G)GG(A/T/G/C)(A/C)G(A/G)CA(A/G)AT(A/T/C)TA-3'	RGRQIY	3-8
6-8	5'- <u>GGGGATCC</u> (A/C)G(A/G)GG(A/T/G/C)(A/C)G(A/G)CA(A/G)AC(A/T/G/C)TA-3'	RGRQTY	3-8
9-11	5'- <u>GGGGATCC</u> AA(A/G)AA(A/G)(C/A)G(A/G)TG(T/C)CC(A/T/G/C)TA-3'	KKRCPY	3-8
12	5'- <u>GGGGATCC</u> AA(A/G)AA(A/G)(C/A)G(A/G)AA(A/G)CC(A/T/G/C)TA-3'	KKRKPY	3-8
13	5'- <u>GGGGATCC</u> AA(A/G)AA(A/G)(C/A)G(A/G)GT(A/T/G/C)CC(A/T/G/C)TA-3'	KKRVPY	3-8
11'	5'-GT(A/T/G/C)TA(T/C)AT(A/T/C)AA(A/G)GA-3'	VYINKE	24-29
12'	5'-AA(C/T)GA(A/G)TT(C/T)AT(A/C/T)AC(A/T/G/C)(A/C)G(A/T/G/C)CA-3'	NEFITRQ	23-29
Hox-specific primer			
1-10	5'- <u>GGGGATCC</u> GA(A/G)(T/C)T(A/G)GA(A/G)AA(A/G)GA(A/G)TT-3'	ELEKEF	15-20
3' primer	5'- <u>GGGTCGAC</u> (T/C)C(G/T)(T/C)C(G/T)(A/G)TT(T/C)TG(A/G)AACCA-3'	RRNQFW	48-52

**FIG. 1.** Primer sequences for PCR. As 5' primers, 10 kinds of mixed oligonucleotide were prepared for group-specific amplification. One more primer was synthesized for amplification of *Hox* genes classified into groups 1 to 10. As the 3' primer, a single primer was used in common. To facilitate analysis of isolated clones, the underlined *Bam*H1 and *Sa*I sites were generated at the 5' ends of primers. The amino acid sequences encoded by these oligonucleotides and their positions in the homeodomain are also indicated.

## MATERIALS AND METHODS

Genomic DNA was isolated from adult medaka fish according to conventional methods (15). Twelve kinds of mixed oligonucleotides, summarized in Fig. 1, were chemically synthesized. A single primer was applied in common as the 3' primer. As the 5' primers, 10 kinds of primer were designed for group-specific amplification and one more primer was used for amplification of *Hox* genes classified into group 1 to 10. PCR was performed as described elsewhere (16). In brief, the 50  $\mu$ l reaction mixture contained 250 ng of genomic DNA, 1  $\mu$ M of primers, 250  $\mu$ M of each dNTP, and 2.5 units of ExTaq polymerase in ExTaq buffer supplied by Takara Co. The mixture was subjected to 30 amplification cycles in Perkin-Elmer/Cetus Thermocycler: 20 sec at 98°C, 1 min at 55°C and 10 sec at 72°C. After fractionation by electrophoresis on 1% agarose gels, the bands of expected size were eluted, and cloned into pT7-Blue vector (Novagen Co.). Nucleotide sequences were determined by the Dye Termination method (Applied Biophysics).

## RESULTS

Since *Hox* genes have been isolated from various animal species and extensively analyzed, it is possible to predict the amino acid sequences encoded by medaka *Hox* genes. Especially amino acid sequences of homeodomains in *Hox* proteins are highly conserved (17). Moreover, regions containing each homeodomain have been shown to be encoded by a single exon without interruption by introns except for the *C. elegans* case. Therefore, PCR methods using medaka genomic DNA should work for amplification of DNA fragments containing homeobox portions as long as appropriate primers are chosen. Six amino acid residues from the 48th to 53rd positions in each homeodomain are WFQNRR without exception. Mixed oligonucleotides whose sequences are complementary to those for WFQNRR were therefore used in common as the 3' primer for PCR. As 5' primers, we first chemically

synthesized 8 kinds of oligonucleotide as summarized in Fig. 1. For example, primer 1 encodes (I/V/L)RTNF (the 4th to 8th positions in the homeodomain) whose sequence is highly conserved among *Hox* genes classified into group 1 but does not exist in *Hox* genes classified into groups 2 to 13. We performed PCR using medaka genomic DNA and primer 1 and the common 3' primer and products were separated by agarose gel electrophoresis. DNA fragments that migrated around 160 bp were extracted from the gel, cloned into the pT7-Blue vector and sequenced. Five clones that had been independently isolated gave the same sequence with each other as indicated by MF-HOX1-1 in Fig. 2. In a similar way, 17 different sequences were revealed as summarized in Fig. 2. Since *Hox* genes classified into group 11 and 12 were not isolated, we synthesized 2 more primers (primer 11 and 12) as indicated in Fig. 1 and performed a PCR survey. MF-HOX12-1 was obtained with this PCR product. The amino acid sequence from the 15th to 20th residues in homeodomains is ELEKEF among *Hox* genes classified into groups 1 to 10 without exception and ELEKEF and WFQNRR have been frequently used as primer sites for amplification of *Hox* genes (18). Therefore, we used this combination of primers and performed further PCR screening. This revealed the presence of a further 9 *Hox* genes in the medaka genome. Although we expected detection of *Hox* genes classified into group 11, neither primers 9-11 nor primer 11 together with the common 3' primer amplified any such *Hox* genes. The nucleotide sequences and the deduced amino acid sequences of the 27 *Hox* genes finally identified in the medaka genome are summarized in Figs. 2 and 3, respectively.

	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47		primer(number of clones)				
GROUP 1	CACCACCAAGCAGCTGACAGAGCTGGAGAAGGAGTTCCATTTCACCAAGTACCTGACTCGGGCTCGCCGTGCGCTGCAATGAGACTCAGGTGAAAATT																																												MF-HOX1-1	1-10(1), 1(5)
																																												MF-HOX1-2	1-10(2)	
GROUP 2	CAACACACAGCTACTGGAAGCTGGAGAAGGAGTTCCACTTCACCAAGTATCTGTGCGCGCGAGGAGGGTGGAAATAGCGGCCCTGTGTGACCTGACCGAAGGACAGGTCAAGGTG																																												MF-HOX2-1	1-10(2), 2-5(3)
	---AC-C-A-A-T-T-C---A---T---T---A-C-C-A---T---G---G---T---G---T---TT-AT-A-G---G																																													

**FIG. 2.** Nucleotide sequences detected in this study which encode a homeodomain portions. The sequences of the regions sandwiched by two primers are indicated. The numbers on the top line correspond to the amino acid positions in the homeodomain. The sequences determined were classified into groups based on the identity of critical amino acids found in each group homeodomain (17). Since the amino acid sequences of the homeodomains classified into 5 to 7 were very similar, they are not sub-classified. The nucleotide sequence in the clone amplified by the group-specific primer is indicated as the reference sequence in each group with bars indicating the same nucleotides. The *Hox* gene classified into group 3 was amplified only by the *Hox*-specific primer. MF-*Hox*-X-1 does not appear to correspond to any group of *Hox* genes. MF-*Hox*13-X was too short to be properly classified. In addition, two *Gbx* genes were identified. On the right, the primers used for isolation of each clone are indicated along with the numbers of clones independently isolated in parentheses. The DDBJ/EMBL/GenBank accession numbers for MF-*Hox* genes determined in this study are AB027023-AB027053.

In addition to these results, four different sequences were revealed as byproducts. It has been known that *Gbx* genes can be amplified by PCR using ELEKEF and WFQNR primers (19). As expected, two *Gbx* genes were identified as indicated in Figs. 2 and 3. Although the amino acid sequence of one clone termed MF-HOX2-66 was very similar to that in a *Hox* gene, it did not appear to be a typical *Hox* gene. Since the Southern hybridization using MF-HOX-X-1 as a probe gave a single sharp band (data not shown), this clone was unlikely to have been an artifact in PCR. Another clone termed MF-HOX13-X was deemed to be an accidental product of primer 1 and the 3' common primer. Since the sequence did not correspond to any of the 27 *Hox* genes described above, it may, however, suggest the presence of another *Hox* gene.

## DISCUSSION

The present search of the medaka genome by PCR identified 27 different *Hox* genes. Since the efficiency of amplification may differ among the target genes, it is likely that we overlooked some *Hox* genes in this PCR survey. In the present study, we used two types of primers, one group specific, and the other *Hox* specific with the exceptions for groups 11 to 13. With the group-specific primers, we could identify 14 *Hox* genes classified into groups 1 to 10 and 4 in groups 12 and 13.

With *Hox*-specific primers, we identified 9 additional *Hox* genes. It is important to emphasize that primers that correspond to ELEKEF and WFQNR worked for

**TABLE 1**

Comparison of the Numbers of *Hox* Genes Classified into 13 Groups among Medaka Fish (MF), Pufferfish (PF), Zebrafish (ZF), and Mammals (MM)

Group	MF	PF	ZF	MM
1	2	2	4	3
2	2	3	2	2
3	1	2	4	3
4	3	3	4	4
5	6	3	4	3
6		2	4	3
7		0	1	2
8	2	2	3	3
9	4	4	5	4
10	3	3	4	3
11	0*	3	5	3
12	1	1	2	2
13	3 (+1)†	3	5	4
sum	27 (+1)	31	47	39

\* The primers used for amplification of *Hox* genes classified into group 11 did not work. However, it is likely that *Hox* genes of group 11 also exist in medaka.

† One more *Hox* gene exists but the sequence identified was too short to be properly classified.

	10	20	30	40	50	60	
ANTP	RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIETIAHALCLTERQIKIWFQNRMRKWKKEN						
GROUP 1	-TK-LT-----	-K----	A--V--	A--Q-N-T-V--			MF-HOX1-1
		--SK----	A--V--	AT-E-N-T-V--			MF-HOX1-2
GROUP 2	NT-L-----	-K--C-P--	V---AL-D----	V-V			MF-HOX2-1
	NN-L-----	-K--C-P--	V---AL-D-S-K-V-V				MF-HOX2-2
GROUP 3		-----C-P--	V-M-NL-N-----				MF-HOX3-1
GROUP 4	-Q-A-----		V---TM--S--V--				MF-HOX4-1
	Q-V-----	Y-----	V---T--S-----				MF-HOX4-2
		-----	T--N-----				MF-HOX4-3
GROUP 5~7	S-----	S---VK----	T-----				MF-HOX5~7-1
	T-----		S-----				MF-HOX5~7-2
		-----	S-----				MF-HOX5~7-3
		-----NH-----	T--N-----				MF-HOX5~7-4
		-----					MF-HOX5~7-5
		-----T--S-----					MF-HOX5~7-6
GROUP 8	S-----	L--P---K--VS---	A---V--				MF-HOX8-1
		--P---K--VS---	G---V--				MF-HOX8-2
GROUP 9	-K-----	L--M---D--Y-V-RV-N----	V--				MF-HOX9-1
		L--M---D--H-V-R--N----	V--				MF-HOX9-2
		LY-M---D--F-V-RI-S-----	V--				MF-HOX9-3
		LY-M---D--L-V-GL-N----	V--				MF-HOX9-4
GROUP 10	-KH-----	L--M---E--L--SKSVH--D--V--					MF-HOX10-1
		L--M---E--L--SRSVH--D--V--					MF-HOX10-2
		L--M---E--L--SRSVN--D--V--					MF-HOX10-3
GROUP 12			--R-LSDR-N-SDQ-V--				MF-HOX12-1
GROUP 13	-QI-LK----	YAASKFI-KDK-RR-SA-TN-S---	VT-				MF-HOX13-1
	-KI-LK----	YAA-KFI-KDK-RK-SAVTN-S---	T-				MF-HOX13-2
	-KV-LK---	R-YAP-KFI-KDK-RR-SAQTN-S---	VT-				MF-HOX13-3
Non classified				T---VT-			MF-HOX13-X
		--SP--C-R--L-M-AG-Q--DQ-V--					MF-HOX-X-1

**FIG. 3.** Amino acid sequences encoded by DNA amplified by PCR. They were deduced from the nucleotide sequences indicated in Fig. 2 with the amino acid sequence of the homeodomain of *Drosophila Antp* used as a reference. Bars indicate the same amino acid as the reference.

amplification of 11 out of the 14 *Hox* genes amplified by group-specific primers. Therefore, we can argue that the majority of *Hox* genes classified into group 1 to 10 were cloned in this study. While we could not reveal the presence of any group 11 *Hox* gene in this study, this does not preclude their existence. Judging from the data for pufferfish and zebrafish (12, 13), it is likely that there are also counterparts in the medaka genome.

Table 1 presents a comparison of numbers of *Hox* genes identified in medaka, pufferfish, zebrafish and mammals (human and the mouse gave the same results). Since assignment of each *Hox* gene into groups is based on the partial sequence of the homeodomain, it may become necessary to change this in future. For example, we classified 6 *Hox* genes into groups 5 to 7. If one of them were to be classified into either group 2 or 3, the numbers of *Hox* genes identified in the present study would be almost identical to those in pufferfish except for group 11. We have now started screening a BAC library of the medaka genome using DNA containing parts of *Hox* genes isolated in this study. However, while present PCR survey can not be regarded as conclusive, the results suggest that the complexity of *Hox* clusters in medaka genome may be more similar to that in pufferfish rather than zebrafish. In addition,

the DNA fragments isolated in the present study should be very useful for identifying BAC clones containing *Hox* genes in the BAC library of medaka DNA.

## ACKNOWLEDGMENTS

We thank A. Koga for helpful discussion. This work was partly supported by Grant No. 09554053 to H.H. from the Ministry of Education, Science, Sports and Culture of Japan.

## REFERENCES

- Ohno, S. (1970) Evolution by Gene Duplication. Springer-Verlag, Heidelberg.
- Schubert, F. R., Nieselt-Struwe, K., and Gruss, P. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 143-147.
- Seimiya, M., Ishiguro, H., Miura, K., Watanabe, Y., and Kurosawa, Y. (1994) *Eur. J. Biochem.* **221**, 219-225.
- Duboule, D., and Dolle, P. (1989) *EMBO J.* **8**, 1497-1505.
- Akam, M. (1989) *Cell* **57**, 347-349.
- Sulston, J., Du, Z., Thomas, K., Wilson, R., et al. (1992) *Nature* **356**, 37-41.
- Martinez, D. E., Bridge, D., Masuda-Nakagawa, L. M., and Cartwright, P. (1998) *Nature* **393**, 748-749.
- Zhang, J., and Nei, M. (1996) *Genetics* **142**, 295-303.
- Scott, M. P. (1992) *Cell* **71**, 551-553.
- Garcia-Fernandez, J., and Holland, P. W. HP. W. (1994) *Nature* **370**, 563-566.



11. Pendleton, J. W., Nagai, B. K., Murtha, M. T., and Ruddle, F. H. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 6300–6304.
12. Aparicio, S., Hawker, K., Cottage, A., Mikawa, Y., Zuo, L., Venkatesh, B., Chen, E., Krumlauf, R., and Brenner, S. (1997) *Nature Genet.* **16**, 79–83.
13. Amores, A., Force, A., Yan, Y., Joly, L., Amemiya, C., Fritz, A., Ho, R. K., Langeland, J., Prince, V., Wang, Y. L., Westerfield, M., Ekker, M., and Postlethwait, J. H. (1998) *Science* **282**, 1711–1714.
14. Postlethwait, J. H., Yan, Y., Gates, M. A., *et al.* (1998) *Nature Genet.* **18**, 345–349.
15. Inagaki, H., Bessho, Y., Koga, A., and Hori, H. (1994) *Gene* **150**, 319–324.
16. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988) *Science* **239**, 487–491.
17. Burglin, T. R. (1994) *in* Guidebook to the Homeobox Genes (Duboule, D., Ed.), pp. 25–71, Oxford Univ. Press, Oxford.
18. Naito, M., Kohara, Y., and Kurosawa, Y. (1992) *Nucleic Acids Res.* **20**, 2967–2969.
19. Matsui, T., Hirai, M., Wakita, M., Hirano, M., and Kurosawa, Y. (1993) *FEBS Lett.* **322**, 181–185.