Proc. Natl. Acad. Sci. USA Vol. 89, pp. 2883–2887, April 1992 Biochemistry

## Hox-1.11 and Hox-4.9 homeobox genes

(*Hox-4.3/Hox-4.2/*homeobox nucleotide sequences)

## ADIL NAZARALI, YONGSOK KIM, AND MARSHALL NIRENBERG\*

Laboratory of Biochemical Genetics, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20892

Contributed by Marshall Nirenberg, December 17, 1991

**ABSTRACT** Mouse Hox-1.11 and Hox-4.9 genes were cloned, and the nucleotide sequences of the homeobox regions were determined. In addition, nucleotide sequence analysis of the homeobox regions of cloned Hox-4.3 and Hox-4.2 genomic DNA revealed some differences in nucleotide sequences and in the deduced homeodomain amino acid sequences compared with the sequences that have been reported.

Homeobox genes code for proteins that bind to specific nucleotide sequences in DNA and either activate or inhibit the expression of the corresponding genes (for reviews, see refs. 1-5). Homeobox proteins are related to one another primarily in the sequence of the 60-amino acid residue DNA-binding-site portion of the protein, the homeodomain. The homeobox family of genes is large; more than 50 mouse homeobox genes or species of cDNA have been reported thus far, and additional homeobox genes undoubtedly will be found in the future. Many homeobox genes reside at neighboring sites in the chromosome in clusters of homeobox genes (5, 6). Whereas the Drosophila genome contains only one copy of the Antennapedia (Antp) and Ultrabithorax (Ubx) clusters of homeobox genes, mammalian genomes contain four copies of the combined Antp-Ubx cluster of homeobox genes, which presumably originated by successive duplications of an ancestral cluster of genes (7). The amino acid sequences of the homeodomains encoded by genes that originated as copies of the same ancestral gene, which are located in different clusters of genes, are more closely related to one another than the homeodomains encoded by other genes within the same cluster. Both the amino acid sequence of the homeodomain encoded by each gene and the order of the genes within the four mammalian Antp-Ubx clusters of genes have been highly conserved during evolution. Why the organization of genes within each cluster has been maintained during evolution is not known, but several clues have been found. There is considerable overlap in the expression of many of the homeobox genes in the Antp-Ubx clusters of genes along the anterior-posterior axis of the embryo, but the anterior border of gene expression is successively displaced towards the posterior, starting with the second gene from the 3' end of the cluster and progressing toward the gene at the 5' end of the cluster (8, 9). Thus, different combinations of homeobox genes are expressed in different regions along the anterior-posterior axis of the embryo (10, 11). In addition, treatment of cultured human embryonal carcinoma cells with retinoic acid results in the gradual, sequential activation of many homeobox genes in each cluster over a period of days, starting with the gene at the 3' end of the cluster and proceeding towards the 5' end of the cluster (6). These results suggest that the order of homeobox genes within each cluster may be involved in determining the topographic position and/or the developmental time of initiation of expression of these homeobox genes in the embryo.

In this report, the nucleotide sequences of the homeobox regions of Hox-1.11 and Hox-4.9 genes are described.<sup>†</sup>

## METHODS AND MATERIALS

Clones of PCR-Amplified Mouse Genomic DNA. The homeobox regions of many mouse homeobox genes were amplified by PCR. Multiple species of oligodeoxynucleotides that correspond to highly conserved sequences in the homeoboxes of many mouse homeobox genes were synthesized with the aid of an Applied Biosystems DNA synthesizer model 380B and purified by OPC (Applied Biosystems) column chromatography. The (+)-oligodeoxynucleotide primers consisted of 64 species of oligodeoxynucleotides, each 28 nucleotide residues long, with a *Sac* I site near the 5' terminus; the (-)-oligonucleotide PCR primers consisted of 48 species of oligodeoxynucleotides, 28 nucleotide residues long, with an *Eco*RI site near the 5' terminus. (See Fig. 2 for the nucleotide sequences of the primers.)

A programmable DNA thermal cycler (Perkin–Elmer/ Cetus) was used for the amplification of DNA. A typical 25- $\mu$ l reaction mixture contained 1  $\mu$ g of BALB/c mouse liver genomic DNA; 50 mM KCl; 10 mM Tris·HCl (pH 8.3); 1.5 mM MgCl<sub>2</sub>; 0.01% gelatin; 15.6 nM of each species of (+)-oligodeoxynucleotide primer and 20.8 nM of each species of (-)-oligodeoxynucleotide primer; 1.0 mM each of dATP, dCTP, dGTP, and dTTP; and 2.5 units of *Taq* polymerase. Reaction mixtures were covered with 50  $\mu$ l of mineral oil and were incubated for 35 PCR cycles; each cycle consisted of incubation for 1 min at 94°C, 2 min at 37°C, and 3 min at 65°C. After the last cycle, the reaction mixtures were incubated for an additional 10 min at 65°C. The DNA was precipitated with ethanol, incubated with *Eco*RI and *Sac* I, and subcloned in pBluescript II KS(+) (Stratagene).

**RNA Probes.** <sup>32</sup>P-labeled ( $\overline{+}$ )-RNA probes were prepared by using a modification of the Stratagene RNA transcription protocol. A typical 10- $\mu$ l reaction mixture contained 40 mM Tris·HCl (pH 8.0), 8 mM MgCl<sub>2</sub>, 2 mM spermidine, 50 mM NaCl, 10 mM dithiothreitol, 10  $\mu$ M [ $\alpha$ -<sup>32</sup>P]UTP (800 Ci/ mmol; 1 Ci = 37 GBq), 0.5 mM ATP, 0.5 mM CTP, 0.5 mM GTP, 194 fmol of linear proteinase K-treated DNA, 10 units of RNase inhibitor, and 4 units of phage T7 RNA polymerase. Reaction mixtures were incubated at 37°C for 30 min; then RNA was precipitated with sodium acetate and ethanol.

Clones of Unamplified Homeobox Genomic DNA. A mouse genomic DNA library in  $\lambda$ GEM-11 (Promega) was screened for some of the homeobox genes that had been found with PCR-amplified DNA. *E. coli* KW251 cells (2 × 10<sup>9</sup>) infected

<sup>\*</sup>To whom reprint requests should be addressed at: Laboratory of Biochemical Genetics, National Heart, Lung, and Blood Institute, Building 36, Room 1C-06, 9000 Rockville Pike, Bethesda, MD 20892.

<sup>&</sup>lt;sup>†</sup>The sequences for *Hox-1.11*, *Hox-4.9*, *Hox-4.3*, and *Hox-4.2* have been deposited in the GenBank data base (accession nos. M87801–M87804, respectively).



FIG. 1. Mouse homeobox gene clusters. Genes that code for proteins with similar homeodomain amino acid sequences that are though to be copies of the same ancestral gene are aligned vertically. The numbers 1 through 13 (from right to left) at the top of the figure represent the vertical sets of related genes in different clusters. Clones of homeobox genes described in this report are shown with shaded backgrounds. Homeobox nucleotide sequences shown in this report are indicated by boxes drawn with thick solid or dashed lines. The chromosomal location of *Hox-X* is uncertain. The boxes drawn with thin dotted lines indicate that no mouse homeobox sequence has been reported; the names of the human *HOX* genes (6, 13) are shown beneath these boxes. Some DNA clones described in this report also are shown beneath the appropriate box; the number of DNA clones found is enclosed within parentheses. DNA clones that begin with P are clones of PCR-amplified mouse genomic DNA; clones prefaced by  $\lambda$  are clones of mouse genomic DNA that were not amplified prior to cloning.

with 25,000 recombinant phages were plated on each 150-mm Petri dish. Phage DNA adsorbed to replica nytran filters (GeneScreenPlus, DuPont) was hybridized overnight at 60°C with <sup>32</sup>P-labeled RNA (35 fmol/ml,  $2 \times 10^6$  cpm/ml) synthesized from cloned PCR-amplified DNA. The hybridization buffer contained 1 M NaCl, 50 mM Tris-HCl (pH 7.6), 1% SDS, and 100  $\mu$ g of yeast tRNA per ml. Filters were washed twice with  $2 \times SSC$  (300 mM sodium chloride/30 mM sodium citrate, pH 7.0) at room temperature for 15 min (each wash), followed by two washes in  $2 \times SSC/1\%$  SDS at 60°C for 60 min (each wash) and finally by one wash in 0.1× SSC at 24°C for 30 min. Filters then were exposed to x-ray film in cassettes at  $-70^{\circ}$ C. Recombinant phage with matching positive signals on autoradiograms of replica filters were cloned. DNA inserts were excised with Sac I and cleaved with various restriction enzymes; some DNA fragments were subcloned into pBluescript II SK(+).

DNA Sequencing. Both strands of cloned DNA fragments were sequenced manually by using Sequenase 2.0 (United

States Biochemical) with universal phage M13 primers or specific primers by the dideoxynucleotide chain-termination method (12), and also by using an automated DNA sequencer (Applied Biosystems Model 373A) with *Taq* DNA polymerase at 70°C, dITP instead of dGTP, dideoxynucleotides or primers labeled with fluorescent dyes, double- or single-stranded DNA preparations, and other components according to the manufacturer's instructions.

## **RESULTS AND DISCUSSION**

**Clones of PCR-Amplified Homeobox DNA.** Small DNA fragments that correspond to part of the homeobox of many mouse homeobox genes were amplified from mouse genomic DNA with the use of sets of primers, each set consisting of multiple species of oligodeoxynucleotides that correspond to conserved nucleotide sequences within the homeobox (nucleotide residues 43–68 and 142–162). The amplified DNA was subcloned, and the chain lengths of the DNA inserts from

				1.5		F.0.C				_1.																								1			1.5	DD (	HEDE				
				(+)	PH III	642																															(*)	P N I		+			
	(+)5'	TCGRC		688	899	CAR	TTT	гөт	TTT	69																									ann	888	GTT	TTA	Sec.	101	TAC	TTRES	et 5'(-)
	(-),	Tuono	cit	G	6	0110 6		C/III	1	-																									nee	6	011	6		6.0	- nc	PERCE	NT BEE
					•	, i	·	•																												•		•		G		HONOL	DGY NOS
		13					60			69	9	75					90					105					120					135		141			150				162	(69-1	418)
Hox	1.11 P8	IB ĠAG	CIC	GAG	AAG	GAA	ΠŤ	CAC	TTC	ARC	FAG	TRĊ	110	TGC	AGA	000	coċ	PGG	GTG	GAA	AIC	500	606	016	CIG	GRT	ΤĪĠ	AÇ C	GRĢ	RGR	CAP	GTĠ	888	GTĠ	TGG	TTC	CAG	ARC	C G G	RG 6	RTĠ	100	
Hox	2.8	GAG	CTĞ	GAG	RAG	GAG	ΠÖ	CRC	TTC	AAT	AAG	TRC	C TG	TGC	CGG	000	€SĪ	CCC	GTC	GAG	AIC	6C <u>T</u>	6C	<b>1</b> 1 G	CIG	GAC	(CTC)	ACC	GAA	AGG	CAC	GTC	AAA	G T 🖸	TGG	11C	CAC	AAC	CGA	1 00 <u>c</u>	ATG	71	[16]
Hox	1.9 PL	25 GAG	CIC	GAG	ARG	GAG	III	CAT	III	BAL	AAG	TAC	CTR	ACT	AGA	GCC	CCH	CGC	AIC	GAC	ATA	000	ARC	IGI	TIN	CRG	CIG	RAT	GRC	ACC	CRG	GIC	AAA	AIC	TGG	IIC	CRG	ARC	000	060	AIG	100	
Hox	1.6	686	CIG	GRG	HRG	686	ΠC	CHC	110	AHC	886	TRC	CIIII	HCH OF CO	CGR	500	CGC	HUU	616	GRG	HIII	600	Line Line	TCC .	CIN	CHU	CIC	HHI	6HG	HCC	CHG	616	HHU	RIC	169	TTC	CHD	HH II	CULC	1060	HIG		[24,25]
Hox	2.9	640	070	COC	HHU	COO	111	CHI	TTO	HHU	HAN	THC	CTC.		CUI	SCC.	000	HUU		0HU COC	HIU	COC		HL L		COC	ere	0.0 T	COIO	HUU ACC	CHD	CTC	NHIL OOC	ATC	TCC	TTC	CHU	NHU	000	000	HIU OTC	67	[10,17]
nux .	4.7	010	i cia	U unu	nnu	0 mm		UNI	161	nnej	n na su	Inc	C (D)	nuu i	Gom	bee	000	000	UIR,	010		0.0	Line	unese a	21.00		τıς	in in i	0n0	nçiol	ene	010	na Qi		100	Π¢	cho	nnç		COL	niu	0.0	[10]
Hox	X P30	GAG	а сто	GAG	886	GRG	TT	CRT	TTC	aec	CGC	TAC	ста	TGC	CGG	000	66C	066	GTG	686	ATG	GCC	880	CTS	CTG	ABC	CIC	AGC	GAG	CGC	CAG	ATC	AAG	ATC	TGG	TTC	CAA	AAC	CGG	CGG	RTG	100	
Hox	1.5	GAG	СТЮ	GAG	AAG	GRG	TTE	CAD	TIC	ARC	CGC	TAC	CTR	ALLIC	CGG	CCG	CGC	CGG	GTG	GRG	ATG	GCC	ARC	CIS	CTG	ABC	CTC	RICIC	GRG	CGC	CRG	ATC	ARG	ATC	TGG	ΠŪ	CAG	AAC	CGG	COC	ATG	95	[24-26]
Hox	4.1	686	; сто	GAG	RAG	GRG	TTC	CRC	TTC	ARC	CGC	T A	C T 🖸	TGC	CGG	CCG	CGC	CGC	GTG	GAG	RTG	GCC	9 A A	C75	CIG	AAC	CIC	ACC	GAÄ	CGC	CAG	ATC	AAG	ATC	T 6 G	τīς	CRG	AAC	CGT	coc	ATG	93	[27]
Hox	2.7	686	с то	GAG	RAG	GAG	τt	CRC	TTC	ABC	CGT	TAI	ШТC	TGC	CGG	000	000	CGG	GTC	GAG	PTG	GCC	AAT	CTG	C⊺G	ANC	CIC	RGC	GAG	CGC	CAG	ATC	<b>A</b> A G	ATC	t G G	110	CAG	AAC	CGI	CGC	ATG	92	[28]
																																		1									
Hox	4.3 P2	E GRG		GAG	AAC	GAA	III	CAT	TIC	ARC	CCT	TAT	CIG	ACC	AGG	AAG	AGG	AGA	ATC	GRG	GIC	TCC	CAT	AC I	CIC	GCC	CIC	RCG	GRG	ACA	CRG	GTR	ARA	RICI	TGG	IIC.	CRA	3 AB	CGG	CGC	ATG	100	
Hox	4.3	686	00	e cull	HHG	GHH	110	C[]]	нű	HHC	CCI	191	C16	нсс	H6 G	HHG	HG G	ACA	нтс	686	GIC	100	CHI	HU	CIE	GÇÇ	ere	HÇ Û	6H 5	HOH	CHG	GIH	ннн	HIC	166	nç	снь	ннε	HPP	Hoff	нть	99	[19]
14					0.00		111				070	707	erc		000	666	CC T	020		600	ore	007	000	000	0.7.0	107	CT C	TOT	000	0.20	cac	97.0	000	от e	TCC	TTO	rac	000	000	200	0.1.0	100	
Hox	4 2 PI	000	ի երե ՅinstP		000	ccal	TIT	ւու	111	aan	800	tet	etc.	acc	acc	000	COT	000	911	000	ete	007	CAC	ann	010	TGT	CIG	In T	GAG	200	rec	ATC	000	ALCI	TEE	TTC	rac	990	000	- Gade	I ATG	100	[15]
104	1.2	0 n <u>n</u>	3	0 0 A M	RHU	0.01		¢υШ		1	1100	011	ciu	nee	100	100		0.00		un	inte		che	nec		101	010	GC I	0110		che	nie	11110		100		che	Inte		LIMA		,,	(13)
Hox	2.9 P3	686	сто	GAR	AAC	GRG	TTT	CRT	ш	BAC	888	780	CTG	ACC	CGT	300	C G G	AGG	616	GRG	ATC	GCC	000	AC C	610	GAG	CIC															100	
Hox	2.9	GAG	стр	GRO (	RAG	GAA	TIT	CRT	TIC	RAC	888	TRC	CIG	AGC	CGT	GCC	CGG	AGG	GTG	GRG	ATC	GCC	GCC	AC C	CIC	GAG	CIC															100	[16,17]
hax	2.9	GAG	а стіс	696	AAG	GRA	117	CPT	Τī	RAC	888	TAC	CTG	AGC	CGT	000	C G G	AGG	GTG	GRG	AIC	GCC	QC C	RCC	CT6	GAG	CIC															96	[18]

FIG. 2. The nucleotide sequences of six homeobox DNA clones, obtained by PCR amplification of mouse genomic DNA, are shown and are compared with the sequences of the most closely related mouse homeobox genes. The numbers at the top correspond to homeobox nucleotide residues. The sequences of oligodeoxynucleotide primers for PCR amplification of DNA are shown at the top. The asterisks above the primers indicate that only one nucleotide residue is present at the position indicated; thus oligodeoxynucleotides will base-pair correctly if the codon sequence in DNA is complimentary to that of the oligodeoxynucleotide but not if the DNA contains other synonym codons for the same amino acid. The symbol  $\dagger$  indicates that (-)-oligodeoxynucleotide primers do not contain A at this position; hence, correct base pairs can form if the DNA contains five of the six arginine codons but not if the DNA contains the arginine codon CGT.

		(+)PRIME	(-)PRIMEF	S PERCEN	T		
	α-f	IELIX 1		CA-HELIX 2	Q-HELIX 3	HOMOLO	G'
		15 21	24	28 38	42 47 52 54	AA BA	SE
Hox	1.11 P88	ĖLEKEFH	FNKYLC	RPRRVEIAALL	DLT ERQUKUUFONR R	100 10	Ō
Hox	2.8	ELEKEFH	FNKYLC	RPARVEIRALL	DLT ERQVKUUFONR R	100 7	1
Hox	4.9 P125	ELEKEFH	FNKYLT	RARRIEIANCL	OLN DTOUK NUFONR R	100 10	a
Hox	1.6	ELEKEFH	FNKYLT	RABRINEIAASL	OLN ETOUK INFONR B	83 7	1
Hox	2.9	ELEKEFH	FNKYLS	RARRUEIAATL	ELN ETOUK INFONR R	75 6	4
Hox	2.9	ELEKEFH	FNKYLS	RABBUELAPTL	ELN ETOUKINFONB BI	75 6	3
Hox	X P30	FLEVEEN	ENRULC	REBRIEMENLI	NIC CONTRINCAND DM	100.10	^
Hax	2.7	FLEKEEH	ENRYLC	REBRIEMANLI	NES ERGERINGOND DE	100 100	2
Hax	1 5	FLEKEFH	ENRVIM	RPRBUEMANLI		100 92	2
Hox	4 1	FIFYFFH	ENRI CUI	ADBBIICMONI I		92 93	2
		LLEKEIT		or novelinite		03 9.	3
Hox	4.3-P24	ELEKEFH	FNPYLT	RKARIEVSHTL	ALT ERQUKINFONR RM	100 100	0
Hox	4,3	ELEKEF	FNPYLT	RKARIEUSHSDL	ALT ERQUKINFOND BO	96 99	9
Hox	4.2-P167	ELEKEFH	FNRYLT	RRABIEIAHTL	CLS EROIKINFONR RM	100 100	٥
Hox	4.2	ELEKEFH	FNRYLT	RARRIEIAHTL	CLP EROIKIUFONR RM	96 99	9
					-		
Hox	2.9-P31	ELEKEFH	FNKYLS	RARRVEIAATL	EL	100 100	0
Hox	2.9	ELEKEFH	FNKYLS	RARRVEIRATL	ELN ETQUKINFONR RM	100 100	Ó
Hox	2.9	ELEKEFH	FNKYLS	RAARVEIAPTL	ELN ETQUKIWFONR RM	96 96	6

FIG. 3. The amino acid sequences deduced from the nucleotide sequences of PCR-amplified cloned mouse genomic DNA (Fig. 2) are shown and are compared with the most closely related mouse homeodomain amino acid sequences. The percent homology between related amino acid sequences for the central region between the PCR primers (amino acid residues 24–47) and the percent homology between nucleotide sequences shown in Fig. 2 (nucleotide residues 69–141) are shown. The numbers at the top refer to homeodomain amino acid residues. The positions of  $\alpha$ -helices 1–3 in the Antennapedia homeodomain (1) of Drosophila also are shown.

93 clones were determined. The nucleotide sequences of some of the DNA inserts also were determined. Of 93 clones examined, 85 were identified as homeobox genes. Two clones, P8 and P91 with different DNA inserts, consist of two homeobox DNA fragments amplified from separate genes that were joined by ligation and cloned. The DNA clones found correspond to 13 mouse homeobox genes as shown in Fig. 1.

The nucleotide sequences of six amplified homeobox genomic DNA clones are shown in Fig. 2 and are compared with the sequences of the most closely related mouse homeobox genes. The percent homology also is shown between the central region of each cloned DNA insert without the primer sequences (homeobox residues 69–141) and the corresponding sequences of the most closely related mouse homeobox genes. The percent homology was calculated for only the central region of the cloned DNA between the primers because oligodeoxynucleotides that hybridize to DNA with some incorrectly paired bases can serve as PCR primers.

Hox 1.1	1 λ33	-89 GROCOTTCCRCCTTCARCTGTATGTGTGTCTCTTGGTTTCCCTTTCCGRGR	-34
		-11 -1 +1	
Hox 1.1	1 λ33	S L E I A D G S G G G S R R L R T A Y T N T Q L	13
Hox 1.1	1 λ33	TCCCTGGAARTAGCTGATGGCAGCGGCGGGGGGGTCCAGGGCGTCTGAGAACCGCGTACACCAACACTCAGCTT	39
Hox 2.8		GGCC-G-T-GC-AATGA-CQC-CA-AC-CGC	
Hox 2.8		6 P G L P E C G - S	
			37
Hox 1.1	1 λ33	TTGGAGCTGGAAAAGGAATTTCATTTCAACAAGTACCTTTGCAGACCCCGCAGGGTGGAAATCGCCGCGCTG	111
Hox 2.8		CGGGCCGG	
			61
10X 1.1	1 233	CTOCHTTTCHCCCHCHCHCHCHCHGTCHHHGTGTGGGTGTGTGCGGGGGGGG	183
HUX 2.0		HC-CHCGGG	
		C K E N Q N S E G K F K N L E D S D K U E E D E	85
Hox 1.1	ι λ33	TGCARGGAGAACCAAAACAGCGAAGGGAAATTTAAAAAACCTGGAGGACTCGGACAAAAGTGGAGGAAGACGAG	255
		EEKSLFEQA	94
Hox 1.1	ι λ33	GRAGAGAGTCACTCTTTGAGCAAGCC	282

Clone Hox-1.11 P8B corresponds to a novel mouse homeobox gene of the Drosophila proboscipedia homeobox class. The nucleotide sequence of clone Hox-1.11 P8B is most closely related to that of the mouse Hox-2.8 gene, but only 71% of the nucleotide residues compared are identical.

Clone Hox-4.9 P125 is a member of the labial class of homeobox genes, probably the first gene in the Hox-4 cluster of homeobox genes (see Fig. 1). The amino acid sequence of the Hox-4.9 homeodomain was reported recently (11), but the nucleotide sequence has not been described. The nucleotide sequence of Hox-4.9 P125 differs considerably from the sequences of other labial class mouse homeobox genes, Hox-1.6 (71% homology) and Hox-2.9 (63-64% homology).

Clone Hox-X P30 may correspond to an unreported murine homeobox gene, probably the third gene in the Hox-3 cluster of homeobox genes (Fig. 1). The nucleotide sequence of the Hox-X P30 homeobox clone is most closely related to that of Hox-1.5 (95% homology).

The nucleotide sequence of clone Hox-4.3 P24 differs from that of Hox-4.3 (14) by only 1 of the 73 residues in the central region between the (+)- and (-)-oligonucleotide homeobox primers, which suggests that clone Hox-4.3 P24 corresponds to the Hox-4.3 gene. The nucleotide sequence of clone Hox-4.2 P167 differs from the sequence reported for Hox-4.2(15) by only 1 nucleotide residue.

Of the 85 homeobox clones obtained, 69 were found to be only 78 nucleotide residues long because of the presence of a Sac I site within the homeobox. Sequence analysis of 9 of the 69 clones revealed one kind of DNA insert with Sac I sites at both the 5' and 3' termini. The nucleotide sequence of a representative clone, Hox-2.9 P31, shown in Fig. 2, is identical to the sequence of the Hox-2.9 gene reported by Rubock et al. (16) and Frohman et al. (17) and differs by only 2 nucleotide residues from the Hox-2.9 sequence reported by Murphy and Hill (18).

The amino acid sequences deduced from the nucleotide sequences of PCR-amplified and cloned DNA are shown in Fig. 3 and are compared with the most closely related mouse homeodomain amino acid sequences. The amino acid sequence of clone Hox-1.11 P8B is the same as that of the Hox-2.8 gene, but the nucleotide sequences differ markedly (71% homology). Clone Hox-4.9 P125 is a *labial* class homeobox gene with an amino acid sequence identical to the sequence recently reported for the Hox-4.9 gene (11). Hox-4.9 P125 differs from the other *labial* class homeobox genes, Hox-1.6 and Hox-2.9, in both nucleotide and amino acid sequences. The amino acid sequences of the homeodomain of clone Hox-4.3 P24 (residues 24–47) differs from that of the Hox-4.3 gene by 1 amino acid residue. Since the corresponding nucleotide sequences differ by only 1 residue, Hox-4.3

FIG. 4. The nucleotide sequence and deduced amino acid sequence of the Hox-1.11 (clone  $\lambda$ 33) homeobox and flanking regions are shown and are compared with the sequences of the most closely related mouse homeobox gene, Hox 2.8 (16). Dashes in the nucleotide sequence or amino acid sequence represent a Hox-2.8 nucleotide or amino acid residue that is identical to the corresponding nucleotide or residue shown for Hox-1.11. The homeobox sequence is enclosed within a box. The arrowhead represents an intronexon junction.



P24 DNA probably corresponds to the Hox-4.3 gene. Similarly, the nucleotide and deduced amino acid sequences of clone Hox-4.2 P167 differ from those of the Hox-4.2 gene by only 1 nucleotide residue and 1 amino acid residue, which suggests that clone Hox-4.2 P167 corresponds to Hox-4.2. The nucleotide and amino acid sequences of clone Hox-2.9 P31 are the same as those of the Hox-2.9 gene (16, 17).

Clones of PCR-amplified DNA also were obtained and sequenced that correspond to *Hox-1.1*, *Hox-1.2*, *Hox-1.3*, *Hox-1.6*, *Hox-2.1*, *Hox-3.4*, and *Hox-4.4* (data not shown).

Homeobox Genomic DNA Clones in  $\lambda$ GEM-11. A mouse genomic DNA library in  $\lambda$ GEM-11 with 15-kilobase (kb) DNA inserts (average size) that were not amplified prior to cloning was screened for homeobox genes with a mixture of <sup>32</sup>P-labeled RNA probes synthesized from PCR-amplified, cloned DNA that correspond to *Hox-1.11*, *Hox-4.9*, *Hox-4.3*, *Hox-4.2*, *Hox-3.4*, *Hox-2.9*, *Hox-1.2*, and *Hox-1.1*. Two million recombinants were screened, and 29 clones of homeobox genomic DNA were obtained. Restriction site analysis revealed seven kinds of DNA inserts that were shown by nucleotide sequence analysis to correspond to seven homeobox genes (*Hox-1.11*, *Hox-4.9*, *Hox-4.4*, *Hox-4.3*, *Hox-4.2*, *Hox-3.4*, and *Hox-1.1*).

Hox-1.11. Two genomic DNA clones,  $\lambda 16$  and  $\lambda 33$ , were found that correspond to Hox-1.11. The nucleotide sequence and deduced amino acid sequence of the Hox-1.11  $\lambda$ 33 homeobox and flanking regions are shown in Fig. 4 and are compared with homeobox sequences of the most closely related homeobox gene, Hox-2.8. The nucleotide sequence of Hox-1.11  $\lambda$ 33 DNA, which was not amplified prior to cloning, was identical to that found with Hox-1.11 P8B cloned from PCR-amplified mouse genomic DNA (nucleotide residues 69-141) shown in Fig. 2. Although only 74% of the Hox-1.11  $\lambda$ 33 and *Hox-2.8* homeobox nucleotide residues are the same, the amino acid sequences of the homeodomains of Hox-1.11 and Hox-2.8 are identical. However, 9 of the 11 Hox-1.11  $\lambda$  33 deduced amino acid residues that precede the homeodomain and 1 amino acid residue after the homeodomain differ from those of Hox-2.8. These results show that Hox-1.11 and

FIG. 5. The nucleotide sequence and deduced amino acid sequence of the homeobox region of the Hox-4.9 gene (clone  $\lambda$ 41) are shown. The amino acid sequence of Hox-4.9  $\lambda$ 41 is compared to the recently reported (11) Hox-4.9 amino acid sequence. The dashes represent Hox-4.9 amino acid residues that are identical to those of Hox-4.9  $\lambda$ 41. The homeobox region is enclosed within a box.

Hox-2.8 are separate genes. The intron-exon junction shown in Fig. 4 at nucleotide residue -36 was identified by comparing the nucleotide sequences of Hox-1.11 genomic DNA and cDNA, which will be described elsewhere (D. Tan, J. Ferrante, A.N., C. Kozak, V. Guo, and M.N., unpublished data); elsewhere we also show that the Hox-1.11 gene resides in mouse chromosome 6, which suggests that the Hox-1.11 gene is a member of the Hox-1 cluster of genes.

The amino acid sequences of the Hox-1.11 and Hox-2.8 homeodomains are identical, which suggests that both species of homeobox proteins may bind to the same or similar nucleotide sequences in DNA. Another pair of homeobox proteins, Hox-1.3 (19, 20) and Hox-2.1 (21–23) also have identical homeodomains.

Hox-4.9  $\lambda$ 41. The nucleotide sequence and deduced amino acid sequence of the homeobox and flanking regions of clone Hox-4.9  $\lambda$ 41 mouse genomic DNA are shown in Fig. 5. The nucleotide sequence of Hox-4.9 has not been reported previously; however, the deduced amino acid sequence ( $\lambda$ 41) is the same as the recently reported Hox-4.9 homeodomain amino acid sequence (11), which suggests that  $\lambda$ 41 DNA is a Hox-4.9 genomic DNA clone. The nucleotide sequence of Hox-4.9  $\lambda$ 41 mouse genomic DNA (cloned from DNA that was not amplified) is identical to that found with PCRamplified mouse genomic DNA (clone Hox-4.9 P125 homeobox nucleotide residues 69–141 shown in Fig. 2).

Hox-4.3  $\lambda$ 40. The nucleotide sequence and deduced amino acid sequence of the homeobox and surrounding regions of Hox-4.3  $\lambda$ 40 DNA are shown in Fig. 6 and are compared with the nucleotide and amino acid sequences of the most closely related homeobox gene, Hox-4.3 (14). The nucleotide sequence of the  $\lambda$ 40 mouse genomic DNA fragment, cloned from DNA that was not amplified, was found to be identical to that of Hox-4.3 P24, derived from PCR-amplified mouse genomic DNA shown in Fig. 2 (homeobox nucleotide residues 69–141). Only 4 of the 209  $\lambda$ 40 nucleotide residues compared differ from those reported for Hox-4.3 (14); however, three of the homeodomain amino acid residues differ from those reported for Hox-4.3. The high nucleotide se-

-10

21

63

45

135

66

198

Hox	4.3	λ40																		-27	TG	TTG	111	TTA	атс	RGCA
Hox	4.3																									
					-1	+1																				
Hox	4.3	λ40	A	Ρ	G	R	R	R	G	R	Q	Ţ	Y	\$	R	F	Q	т	L	F	L	Е	κ	Е	F	L
Hox	4.3	λ40	60	тсс	TGG	TAG	ACG	GAG	AGG	AAG	ACA	AAC	CTF	ICAG	TCG	CTT	CCR	AAC	CC1	'AGA	IGTT	GGA	AAA	GGA	ATT	CCII
Hox	4.3																	0	66G-							
Hox	4.3		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	R	Ų	-	-	-	-	-	-	-
						_												<b></b>	_							
			F	Ν	Ρ	Y	L	T	R	ĸ	R	R	I.	ε	U	S	Н	T	L	A	L	T	E	R	Q	V
Hox	4.3	λ4D	TT	TAR	CCC.	TTA	TCT	GAC	CAG	GAA	IGAG	GAG	AAT	CGP	IGGT	CIC	CCR	TAC	:TC1	GGC	CCT	CAC	GGA	GAG	ACA	GGTA
Hox	4.3																	0	;							
			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	s		-	-	-	-	-	-	-
			ĸ	1	μ	F	Q	N	R	R	M	κ	u	ĸ	ĸ	Ε	Ν	М	K	D	Κ	F	Ρ			
Hox	4.3	λ40	AA	ART	CTG	GTT	CCA	GAA	CAC	GAC	SAAT	GAP	ATC	GAP	IARF	AGGA	GAA	CIAF	ICAF	AGA	ICRA	GTT	TCC	Ţ		
Hox	4.3																							-		
			-	-	-		-	-	-	+-	-	-	-	-	-	-	-	-	-	-	-		-			

FIG. 6. The nucleotide sequence and deduced amino acid sequence of the homeobox and surrounding regions of  $Hox-4.3 \lambda 40$ mouse genomic DNA are shown and are compared with the Hox-4.3 nucleotide and amino acid sequences reported (14). Only nucleotide and amino acid residues of Hox4.3 that differ from those of clone  $\lambda 40$  are shown; residues that are the same are indicated by dashes. The homeobox region is enclosed within a large box. Nucleotide and amino acid residues that differ are enclosed within small boxes. The arrowhead represents an intron-exon junction reported for Hox-4.3 (14).

		HONOLOGY
	1 10 20 30 40 50 60	AA BASE
Hox 1.11 λ33	SRRLRTRYTNTOLLELEKEFHFNKYLCRPRRUEIARLLDLTEROUKUNFONRRMKHKROT	100 100
Hax 2.8		100 74
HOX IK HUMAN		100 90
Hox 4.9 241	PSRIRTHFSTKOLTELEKEFHFNKYLTRARRIEIANCLOLNDTOVKINFONRAKKKKRE	100 100
Hox 4.9		100
HOX 46 HUMAN	SHHH	97
Hox 1.6	-N-UEEEEE	88 77
Hox X P30	ELEKEFHENRYLCHPRRVEMANLLNLSEROIKINFONRAM	100 100
Hox 2.7		100 92
Hox 4.3 λ4D	REAR THE REPORT OF THE REPORT	100 100
Hox 4.3		95 98
	and the second	
Hox 4.2 λ6	PKRSRTRYTRQQVLELEKEFHENRYLTRRRRIEIAHTLCLSERQIKIUFQNRRMKUKKDH	100 100
Hox 4.2	p	98 99

PERCENT

FIG. 7. The amino acid sequences of the homeodomains encoded by five mouse homeobox genes deduced from the nucleotide sequences of cloned DNA are shown and are compared with the amino acid sequences of the most closely related mouse or human homeodomains. The percent homology between the amino acid sequences of related homeodomains and between the corresponding homeobox nucleotide sequences also are shown. Only differences in amino acid sequence are shown. Dashes represent identical amino acid residues.

quence homology between  $\lambda 40$  and Hox-4.3 suggests that clone  $\lambda 40$  DNA corresponds to the Hox-4.3 gene. However, the possibility that clone  $\lambda 40$  DNA corresponds to a novel homeobox gene, Hox-1.12, the eighth gene in the Hox-1cluster of homeobox genes shown in Fig. 1, is not ruled out.

A summary of results is shown in Fig. 7. Homeodomain amino acid sequences deduced from the nucleotide sequences of cloned mouse genomic DNA are shown and are compared with the most closely related sequences reported for mouse or human homeobox proteins. The amino acid and nucleotide sequence homologies also are shown. The amino acid sequence of the Hox-1.11 homeodomain is identical to that of Hox-2.8; however, 9 of the 11 amino acid residues before the homeodomain and 1 amino acid residue after the homeodomain differ from those of Hox-2.8. Many differences also were observed in the nucleotide sequences of the *Hox-1.11* and *Hox-2.8* homeobox regions. The mouse Hox-1.11 homeodomain is the equivalent of the recently reported human HOX-1K homeodomain (6).

The amino acid sequence of the Hox-4.9  $\lambda$ 41 homeodomain is identical to the recently reported Hox-4.9 homeodomain amino acid sequence (11). The mouse Hox-4.9 homeodomain is the equivalent of the human HOX-4G homeodomain (13).

Six of the 73 Hox-X P30 nucleotide residues differ from the corresponding sequence of the most closely related homeobox gene, Hox-2.7; however, the deduced amino acid sequence of Hox-X P30 is the same as that of Hox-2.7. The cumulative error due to misincorporation of bases during DNA amplification was estimated by comparing the DNA sequences of 10 clones of mouse genomic DNA subjected to 35 cycles of DNA amplification (730 nucleotide residues compared) with mouse genomic DNA sequences that were not amplified prior to cloning, which correspond to Hox-1.11, -4.9, -1.1, -3.4, -4.2, -4.3, and -4.4. No DNA amplification errors were detected. Comparison of the nucleotide sequences of 13 additional clones of PCR-amplified DNA that correspond to Hox-1.2, -1.3, -1.6, -2.1, and -2.9 with published nucleotide sequences revealed only 2 residues that differ (906 nucleotide residues compared). Hence, no more than 2 misincorporated nucleotide residues were found per 1636 residues sequenced in DNA molecules that were cloned after 35 cycles of DNA amplification; that is, the error due to PCR amplification of DNA is no more than 1 residue per 818 nucleotide residues of cloned DNA. Thus, we think it more

likely that *Hox-X* P30 DNA corresponds to a homeobox gene that has not been reported previously, such as the third gene of the *Hox-3* cluster of homeobox genes shown in Fig. 1, rather than a DNA clone with an erroneous sequence due to misincorporation of 6 nucleotide residues during DNA amplification. However, the nucleotide sequence of a *Hox-X* mouse genomic DNA clone that was not amplified prior to cloning is needed to confirm the nucleotide sequence of *Hox-X* P30.

Also, as shown in Fig. 7, the deduced amino acid sequences of the Hox-4.3  $\lambda$ 40 and Hox-4.2  $\lambda$ 6 homeodomains were found to differ from the reported sequences (14, 15) by 3 and 1 amino acid residues, respectively.

- Gehring, W. J., Müller, M., Affolter, M., Percival-Smith, A., Billeter, M., Qian, Y. Q., Otting, G. & Wüthrich, K. (1990) *Trends Genet.* 6, 323-329.
- 2. Dessain, S. & McGinnis, W. (1991) Curr. Opin. Genet. Dev. 1, 275-282.
- 3. Laughon, A. (1991) Biochemistry 30, 11357-11367.
- 4. Scott, M. P., Tamkun, J. W. & Hartzell, G. W., III (1989) Biochim. Biophys. Acta Rev. Cancer 989, 25-48.
- 5. Kessel, M. & Gruss, P. (1990) Science 249, 374-379.
- Simeone, A., Acampora, D., Nigro, V., Faiella, A., D'Esposito, M., Stornaiuolo, A., Mavilio, F. & Boncinelli, E. (1991) Mech. Dev. 33, 215-228.
- 7. Kappen, C., Schughart, K. & Ruddle, F. H. (1989) Proc. Natl. Acad. Sci. USA 86, 5459-5463.
- Graham, A., Papalopulu, N. & Krumlauf, R. (1989) Cell 57, 367–378.
- 9. Duboule, D. & Dolle, P. (1989) EMBO J. 8, 1497-1505.
- 10. Lewis, E. B. (1978) Nature (London) 276, 565-570.
- 11. Hunt, P., Gulisano, M., Cook, M., Sham, M.-F., Faiella, A., Wilkinson, D., Boncinelli, E. & Krumlauf, R. (1991) Nature (London) 353, 861-864.
- 12. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- Acampora, D., D'Esposito, M., Faiella, A., Pannese, M., Migliaccio, E., Morelli, F., Stornaiuolo, A., Nigro, V., Simeone, A. & Boncinelli, E. (1989) Nucleic Acids Res. 17, 10385– 10402.
- Izpisua-Belmonte, J. C., Dolle, P., Renucci, A., Zappavigna, V., Falkenstein, H. & Duboule, D. (1990) Development 110, 733-745.
- Featherstone, M. S., Baron, A., Gaunt, S. J., Mattei, M.-G. & Duboule, D. (1988) Proc. Natl. Acad. Sci. USA 85, 4760–4764.
- Rubock, M. J., Larin, Z., Cook, M., Papalopulu, N., Krumlauf, R. & Lehrach, H. (1990) Proc. Natl. Acad. Sci. USA 87, 4751-4755.
- 17. Frohman, M. A., Boyle, M. & Martin, G. R. (1990) Development 110, 589-607.
- 18. Murphy, P. & Hill, R. E. (1991) Development 111, 61-74.
- Odenwald, W. F., Taylor, C. F., Palmer-Hill, F. J., Friedrich, V., Jr., Tani, M. & Lazzarini, R. A. (1987) *Genes Dev.* 1, 482-496.
- Fibi, M., Zink, B., Kessel, M., Colberg-Poley, A. M., Labeit, S., Lehrach, H. & Gruss, P. (1988) Development 102, 349–359.
- Hauser, C. A., Joyner, A. L., Klein, R. D., Learned, T. K., Martin, G. R. & Tjian, R. (1985) Cell 43, 19-28.
- Jackson, I. J., Schofield, P. & Hogan, B. (1985) Nature (London) 317, 745-748.
- Krumlauf, R., Holland, P. W. H., McVey, J. H. & Hogan, B. L. M. (1987) Development 99, 603-617.
- Baron, A., Featherstone, M. S., Hill, R. E., Hall, A., Galliot, B. & Duboule, D. (1987) *EMBO J.* 6, 2977–2986.
- LaRosa, G. J. & Gudas, L. J. (1988) Mol. Cell. Biol. 8, 3906– 3917.
- McGinnis, W., Hart, C. P., Gehring, W. J. & Ruddle, F. H. (1984) Cell 38, 675–680.
- Lonai, P., Arman, E., Czosnek, H., Ruddle, F. H. & Blatt, C. (1987) DNA 6, 409-418.
- Graham, A., Papalopulu, N., Lorimer, J., McVey, J. H., Tuddenham, E. G. D. & Krumlauf, R. (1988) Genes Dev. 2, 1424–1438.