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.[†]

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- μ l reaction mixture contained 1 μ g of BALB/c mouse liver genomic DNA; 50 mM KCl; 10 mM Tris•HCl (pH 8.3); 1.5 mM MgCl₂; 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 μ 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. ³²P-labeled (+)-RNA probes were prepared by using a modification of the Stratagene RNA transcription protocol. A typical 10- μ l reaction mixture contained 40 mM Tris·HCl (pH 8.0), 8 mM MgCl₂, 2 mM spermidine, 50 mM NaCl, 10 mM dithiothreitol, 10 μ M [α -³²P]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 λ GEM-11 (Promega) was screened for some of the homeobox genes that had been found with PCR-amplified DNA. *E. coli* KW251 cells (2 × 10⁹) infected

^{*}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.

⁺The sequences for *Hox-1.11*, *Hox-4.9*, *Hox-4.3*, and *Hox-4.2* have here dependent in the Conflort data here (according to a 197991) M87804, respectively).

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



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 λ 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 ³²P-labeled RNA (35 fmol/ml, 2×10^6 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 μ 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° 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

										- T -																																	
				(+)	PBI	1ERS																														(.) PR	I MB	RS				
																																						•	•	t			
	(+)5'	TGGR	GCT	C GPA	888	GAR	Ш	CAT	III	AA																								AC	C PA	A GT	T TT	A G	00 10	T TA	C ITF	AGCT	5'(-)
				G	i Q	Ģ	¢	C	С																											6	ç	U	Ų	C C	PEF	ICEN 1	REF.
							60					76										1.06					1.20					125	1.4	,		10				0 14	1 (4)		Y NUS.
Hov	11 P	98 ¹ 56	c ct	e cer			TTT	nan	TIC		car.	TOC	CT7	TEC	606	cee.	000	209	6TG	699	атг	100	606	076	E 16	Get	120	асг	696	868	C 6 6	133	רו נה פפו	έte	S II	10 n ne	с ee	r r	aa aa	01 Ta D:	2 (0)	17 7 10	0)
Hox	2.8	50 SR SR	G ET	GIGAD	886	i GAG	нid	CRC	TIC	eelik	AHG	TRC	citici	TGC	1000	COD	COT	ΠώČ	61Č	GHG	ATC	GCT	600	m G	CTG	GRE	ແກ່ບົ	RCC	688	RGIC	Celtil	GTCI	88 61	ο'n	6 TT-	с ся	G RR		680 RC	E AT	6 7	n i	[16]
						au				٦			œ		30-63												00												· @·				
Hox	4.9 PL	25 GA	6 C!	C GAC	AAC	G A G	TT	CAT	TTT	RAT	AAG	TRC	CTR	AC T	AGA	GCC	CGA	CGC	AT C	GAG	ATR	600	AAC	TGT	ITA	CAC	CTG	AAT	GAC	90A	CRG	GT <u>C</u> A	AA AT	010	GIT	C CA	G AR	сε	6 <u>6</u> CC	C RT	6 16	0	
Hox	1.6	GR	G CT	GIGAC	A A G	G G A G	TTC	CAC	1 IC	ANC	RAG	TAC	CIL	AC(A)	CIGA	ទ៩ច្រ	CCC	RIGIC	GTIS	GAC	ATT	6C C	CCC .	TCC	CTA	CAG	CIC	AAT	GAG	ACC	CAG	GTG	IAIC RT	C	6 IT	C CA	C AA	T] C	CIC CC	C AT	6 1	n [24,25]
Hox	2.9	GA	6 CT	GGAC	AAC	GAA	III	CAT	TTC	ARC	R BAA	TAC	CTG	ACC.	CEL	SEC	CGG	ACC	GILC	GAG	ATC	000 000	GCC	ACC.	CTG	GAG	CTC	RAT	GRIA	ACC.	CAG	GTG	IAIG AT	CITC	G IT	C CA	G AR	с с 2	66 00	C AT	с е с	4 [16,17]
Hox	2.9	68	5 U I	<u>ы</u> она	HHU	5 UHH		CHI	1 IU	HHU	HH	THC	сıы	HOU	цчш	500	Cob	Hon	UI U	0.40	ню	60 <u>0</u>	ugu	HUC	L D	ынь	μg	HH I	UNH	нцы	снь	51511	нын	en	ι II	с сн	ь нн	c c	66 CC	IC HI	υι	1)	[18]
Hox	X P30	GA	6 .01	0.090		0.80	TI	CAT	TIC	aac	CGC	TRC	стя	TGC	000	000	233	066	516	686	AT G	000	986	CIS	стс	880	CIC	AGC	GRG	CGC	CAG	87C (IAG AT	с то	G II	C C8	8 88	сс	66 00	G AT	G 10	10	
Hox	1.5	GA	G CT	C GAC	680	GAG	TTC	CRC	TIC	AAC	CGC	TAC	CTA	FILE	CGG	CCG	CGC	CGG	GTG	GRG	ATG	GCC	ARC	CIS	CTG	ABC	CTC	ACIC	GRG	CGC	CAG	ATC (AG AT	C TO	GIT	I CA	G AA	c c	GG CC	AT D	6	5 1	24-26]
Hox	4.1	6A	G CT	G GAO	S RAC	GAG	TTC	CRC	TTC	AAC	CGC	TAT	CTC	TGC	CGG	CCG	CGC	CGC	GTG	GAG	ATC	GCC	ARC	CTG	CTG	AAC	CTC	ACC	GAA	¢GĊ	CAG	ATC (IAG A1	CIT	G TT	Ĉ CA	G AA	сс	GTCC	ic at	G	3	[27]
Hox	2.7	6A	G CT	G GAC	6 890	G G A G	τt	CRIC	TIC	AAC	CGT	TAT	ITC	100	000	000	C G C	CGG	GTC	GAG	PIC	GCC	AAT	CTG	CIC	AUC	CIC	RGC	GRG	CCC	CAG	ATC 1	IAG A1	010	GIT	с ся	ig aa	с с	GL CC	杞 AT	6 9	2	[28]
Hox	4.3 P2	4 GA	6 01	C 680	5 RAC	5 GAA	111	CRI	110	AHIC	001	101	CIG	RCC	ACC	RAG	AC C	ACA	HIC	GRG	610	100	CHI	HC 1	CIG OTC	GCC	010	HCG	GHG	HCH.	CHG	GIH J	IHH H1	CIN	GII	C CH	H HH	C C	66 CC 266 734	-C HI -70 of	6 10	10	r
нох		ųн	¢Шι	C OHE	a und	, 0HH	16	պո	i u	nne.	UL I	101	610	HUU	HUU	нно	nu u	non	nec	010	erc	100	CHI	ng.	CIU	566	cic	nço	ons	non	CHO	01111		1	6 F)	c cn	£ nn	¢ [fi	իսնե	@ 01	0 3		f (a)
Нох	4.2 P1	67 GB	6 CT	C 686	9 880	5 686	TTT	080	ш	RAC	856	TRT	CIG	800	866	000	CGT	066	RTT	688	810	667	CRC	800	ета	TGT	CTG	TOT	GAG	000	CAG	ATC 1	IAG AJ	ic te	GIT	с ся	G AA	сс	GG CC	ас ат	'G 10	10	
Нох	4.2	GA	CT	G GAR	A RAC	GRA	TIT	¢яЛ	111	ARC	AGG	TRT	CIG	ACC	AGG	CGC	CGT	CSG	AIT	GRA	RTC	GC T	CAC	AC C	CTG	TGT	CIG	[CCT	GAG	CGC	CAG	ATC 1	ING AT	CIT	GIT	C CA	G AA	сс	GG 🗐	🗿 AT	6	99	[15]
			-			-																						-															
Hox	2.9 P3	I GA	G CT	C GAR	A NAC	GRG	TIT	CRT	TTT	RAIC	898	180	CTG	RGC	CGT	000	CGG	AGG	GTG	GRG	RIC	000	GCC	RC C	610	GAG	CIC							1							10	10	
Hax	2.9	GA	G CT	G GAC	RAC	5 GAA	ш	CRT	TIC	RAC	ARA	THC	CIG	RGC	CGT	000	CGG	RGG	010	GRG	RIC	GCC	000	RCC	013	GAG	010							i							10	10	16,17]
Pox	< 2.9	GR	6 01	EI 6.9(0	A HUC	5 68 <u>8</u>	111	CPT	ΠC	нис	ння	THC	UIG	HGC	ÇGT	600	000	HGG	616	686	ніс	6C[j]	gee	нсс	016	UHG	CIC.															10	[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 DNA contains five of the six arginine codons but not if the DNA contains the arginine codon CGT.

		(+)PRIME	IS		č	-)PRIME	IS PERI	CENT
	α-f	ELIX 1		CA-HELIX 2	α-HEL	18 3	HOM	<u>ology</u>
	1	15 21	24 2	28 38	42 47	52 54	AA I	BASE
Hox	1.11 P80	ÉLEKEFH	FNKYLC	RPRRVEIRALL	DLT ĖRQUKŮ	WFONR RI	100	100
Hox	2.8	ELEKEFH	FNKYLC	RPARVEIAALL	DLT ERQUKU	WFONR RI	1 100	71
Hox	4.9 P125	ELEKEFH	FNKYLT	RARRIEIANCL	OLN DTOUKI	WFOHR R	1 100	100
Кох	1.6	ELEKEFH	FNKYLT	RABBUELAASL	OLN ETOUKI	WFONR BI	1 83	71
Hox	2.9	ELEKEFH	FNKYLS	BARRUE I ARITL	ELN ETOUK	UFONR R	1 75	64
Hox	2.9	ELEKEFH	FNKYLS	RARRUEIAPTL	ELH ETQUKI	UFONB BI	1 75	63
Hax	X P30	FLEKEEH	ENBYLC	BPBBUEMANI I	NIS FROIKI	UFONB B	1 100	100
Hox	2.7	ELEKEFH	FNRYLC	RPRRUEMANLL	NLS EROIK	UFONR R	1 100	92
Hox	1.5	ELEKEFH	FNRYLIN	RPRAVEMANLL	NUT EROIK	HEONE B	1 92	95
Hox	4.1	ELEKEFH	FNALCU	APRAVEMANLL	HLT ERQIKI	HEONA AI	1 83	93
Hox	4.3-P24	ELEKEFH	ENPYLT	RKRALEVSHTL	ALT EROUKI	I WFONR R	1 100	100
Hox	4.3	ELEKEF	FNPYLT	RKARIEUSHSJL	ALT ERQUKI	WFONR RI	1 96	99
Hax	4 2-P167	EL EKEEH			CLS EBUIKI		1 100	100
Hox	4.2	ELEKEFH	FNRYLT	RRRRIEIAHTL	CLP ERQIKI	WEONE RI	1 96	99
Hox	2.9-P31	ELEKEFH	FNKYLS	RARRVEIAATL	EL		100	100
Hox	2.9	ELEKEFH	FNKYLS	RARRVEIRATL	ELN ETQVKI	HEONB BI	1 100	100
Hox	2.9	ELEKEFH	FNKYLS	RARRVEIAPTL	ELN ETQUK	WEONR R	1 96	96

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 α -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.

2885

Proc. Natl. Acad. Sci. USA 89 (1992)

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 sequence 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

Hox 1.11	λ33				- 85	, G	ACC	стт	ĊCA	сст	TCAI	ACT	GTA	TGT	GTG	TCT	CTT	GTT	ĠĠŦ	TTC	сст	TTC	TGC	AGAA	-34
		-11									-1	+1]
Hox 1.11	λ33	S	L	E I	Ĥ	D	G	S	G	G	G	s	R	R	L	R	T	A	Y	Т	Ν	T	Q	L	13
Hox 1.11 2	λ33	TCC	CTG	GAAA	TAGO	CTGA	TGG	CAG	CGG	CGG	GGGG	нtс	CAG	GCG	TCT	GAG	AAC	CGC	GTA	CAC	CAA	CAC	TCA	GCTT	39
Hox 2.8		66-	-cc	- G - T	~GC-	-ค	RT-	~G-		-A-	CI	d	-C-	CA-	A	-C-	c	G	C				G	AG	
Hox 2.8		G	Р	GL	Ρ	Ε	C	G	-	S	~	-	-	-	-	-	-	-	-	-	-	-	-	-	l
		L	E	LĒ	ĸ	E	F	Н	F	N	ĸ	Ļ	L	С	R	Ρ	R	R	V	E	T	A	A	L	37
Hox 1.11	λ33	TTG	GAG	стбб	AAAA	AGGA	ATT	TCA	TTT	CRF	ICAA	GTA	сст	TTG	CAG	ACC	CCG	CAG	GGT	GGA	AAT	CGC	CGC	GCTG	111
Hox 2.8		¢			-G		G	C	C		T			G	- C -	G	G~-	TC-	C	C	G		T	CT	
		-	-		-		-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	l
		Ι.			_							-				_						~	-	6	
			U	LI	E	н	Ų	U LOCT	K	U	H	+ 	Ų TOO	N COO	H	K COC	П 00Т	K	H	К 700	к СОС	ų con	1	for the second	103
Hox I.II.	133	1010	онг		LLUI	10110	HUH C	HUI	онн	MUI	610	611	I LH	ынн	ιιυ	000	нн I 0	ынн	OLH	1111	OHU	61 H	nnu r	de no	103
HDX 2.0			C	ι-ι-		-H		0	(-ţ		ι			нц -	U		н	U	-L-		u	GU	
		[-				_	_						_				-	_	_	_	_		5	
		С	ĸ	E١	Q	н	s	Ε	G	κ	F	ĸ	И	L	ε	D	S	D	κ	Ų	Ε	٤	D	Ε	85
Hox 1.11	λ33	TGC	AAG	GRGA	ACCA	ARAF	CAG	CGA	AGG	GAF	ATT	TAA	AAA	CCT	GGA	GGA	стс	GGA	CAR	AGT	GGA	GGA	AGA	CGAG	255
		£	E	ĸs	L	F	E	Q	ß																94
Hox 1,11	λ33	GRA	GAG	AAGT	CHC.	тстт	TGF	IGCA	AGC	C															282
				-																					

FIG. 4. The nucleotide sequence and deduced amino acid sequence of the Hox-1.11 (clone λ 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.

```
-10
Hox 4.9 241 -51
                        ΚL
                            SEYGATSP
                                               SRI
                                                      R
Hox 4.9 X41 CCTTGTCTTTATGTTGCAGGCAARCTGTCCGRATATGGAGCCACAAGCCCTCCCAGTGCCATCCGCACAART
                                                              21
Hox 4.9
                                                SAIRTN
         F S T K Q L T E L E K E F H F N K Y L T R R R R
                                                              31
Hox 4.9 241
         93
Hox 4.9
          STKQLTELEKEFHFNKYLTRARR
           E I A N C L Q L N D T Q U K I W F Q N R R M K
                                                              55
         ATCORGATAGCCARCTGTTTACAGCTGRATGRCACCCAGGTCARRATCTGGTTCCAGARCCGTAGGATGARG
Hox 4.9 241
                                                              165
Hox 4.9
          E I A N C L D L N D T O U K I U F O N R R M K
         Q K K R E R E G L L A T A A S V A S I K L P R S
                                                              79
Hox 4.9 241
         CAGAAGAAGAAGAGGGAACGAGAGGGGGCTTCTGGCCACAGCTGCCTCTGTGGCCTCGATTARGCTTCCCCGGTCA 237
Hox 4.9
         OKKRE
           T S P
                  IKSGRNLGSPSQRQEPS*
                                                              100
Hox 4.9 241 GRARCHAGTCCCATCAARTCT6GCC6GRATCTA6GAA6CCCTTCTCA6GCTCAA6A6CCTTCCT6A
                                                              303
```

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 λ GEM-11. A mouse genomic DNA library in λ 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 ³²P-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 λ 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 λ 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 λ 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 λ 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 λ 41) are shown. The amino acid sequence of Hox-4.9 λ 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 λ 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 λ 41. The nucleotide sequence and deduced amino acid sequence of the homeobox and flanking regions of clone Hox-4.9 λ 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 (λ 41) is the same as the recently reported Hox-4.9 homeodomain amino acid sequence (11), which suggests that λ 41 DNA is a Hox-4.9 genomic DNA clone. The nucleotide sequence of Hox-4.9 λ 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 λ 40. The nucleotide sequence and deduced amino acid sequence of the homeobox and surrounding regions of Hox-4.3 λ 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 λ 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-

					- 1	+1																				
x 4.	3	λ40	A	Ρ	G	R	R	R	G	R	Q	Ţ	Y	5	R	F	Q	Т	L	F	L	Е	κ	Е	F	L
× 4,	3	λ40	GCI	100	TGG	AG	ACG	GAG	AGG	AAG	AC AI	AAC	СТА	CAG	TCG	стт	CCRI	aaci	ссти	a G A I	GIT	GGA	AAA	GGA	ATT	CCTT
x 4.	3																	G I	GG							
x 4.	3		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	R_	Ų	-	-	-	-	-	-	-
			-							~		D		r		c		F	٦.	0		Ŧ	r	n	•	
	_			п - п	۲ ۵۵۵	Y	L 707		n	С.О.О.	н сос	н 000	1	E oco		5 	n n	Ľ.,	Ľ.,	п ссп.	L 007	1	с. сс. п	п сос	V DOD	0
ох ¶.	3	74N	111	нн	uuu	н	101	GHU	СНЬ	ынн	6H6	ьнъ	BHI	ССН	661	cit	ссн	IHC.	μu	06L		ιнι	66H	бно	нсн	001H
1 X 4	3																	6								
			-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	S		-	-	-	-	-	-	-
						r	•	5.1			м	r		v	~	c.	ы		~	n	r	r	D			
			1 .		м	۲	Ų	n	ĸ	К	н	ĸ	W	ĸ	ĸ	E	n	11	N.	U	ĸ	r	r			

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 Hox 4.3 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 reprenon junedon reported Hox-4.3 (14).

						HOH	DLOGY
	1 10	20	30	10	50 60	88	BASE
Hox 1.11 λ33	SRRLRTAYTHTQ	LLELEKEFHFNK	YLCRPRRUEIF	ALLDÍTERQU	кииғо́нввякнквоі	100	100
Hox 2.8						100	74
ROX IK HUMAN						100	90
Hox 4.9 241	PSRINTNESTKO	LTELEKETHENK	YLTRARRIEIF	нссоінотос	KINFÓNBRUKOKKRÉ	100	100
Hox 4.9						100	
HOX 45 HUMAN	Ş			H-··		. 97	
Hox 1.6	-N-UT		V	-ASE		- 88	77
Hox X P30		ELEKEFHFNR	VLCAPBRUEMA		ктиео́мвал	100	100
Hox 2.7						100	92
Hox 4.3 λ40	BBBGBDTYSEF	TI ELEKEÉLENR				1 100	100
Hox 4.3		-RU				. 95	98
HOX 4.2 AD	PERSETRATEO	IVLELEKEFHFNR	YLIBBER	HITLCLSERQI	KINEONBRUKNKOF	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 λ 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 λ 40 and Hox-4.2 λ 6 homeodomains were found to differ from the reported sequences (14, 15) by 3 and 1 amino acid residues, respectively.

- 1. 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.
- 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.
- 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.
- 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., Schoffeld, P. & Hogan, B. (1985) Nature (London) 317, 745–748.
- 23. Krumlauf, R., Holland, P. W. H., McVey, J. H. & Hogan, B. L. M. (1987) Development 99, 603-617.
- Baron, A., Fcatherstone, 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.