Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 10706–10710, December 1991
Developmental Biology

Identification of 10 murine homeobox genes

(degenerate oligonucleotide screening/mouse development/homeodomain sequence comparison)

GURPARKASH SINGH*†, SATBIR KAUR*, JEFFREY L. STOCK*, NANCY A. JENKINS‡, DEBRA J. GILBERT‡, NEAL G. COPELAND‡, AND S. STEVEN POTTER*

*Department of Basic Sciences, Children's Hospital Medical Center, Elland and Bethesda Avenues, Cincinnati, OH 45229, ‡Mammalian Genetics Laboratory, ABL–Basic Research Program National Cancer Institute–Frederick Cancer Research and Development Center, Frederick, MD 21702

Communicated by Frank H. Ruddle, August 7, 1991

ABSTRACT **In *Drosophila* a number of genes important in establishing segmentation patterns and in determining segment identities have been shown to carry the homeobox sequence. Over 30 murine homeobox genes have been cloned, many on the basis of sequence homology to *Drosophila* prototypes. Here we report the cloning and sequencing of 10 new and 6 previously known homeobox genes by screening a murine genomic library with a 768-fold degenerate oligonucleotide corresponding to the most conserved 8-amino acid motif in the recognition helix of the homeodomain. Eight of these new homeobox genes have been chromosomally mapped. Four genes do not belong to any of the known homeobox gene clusters but instead map to new locations on chromosome 1 (single gene) and chromosome 5 (three genes). Sequence comparisons indicate that two of these are very closely related and represent a distinct new category of homeobox genes. The remaining four mapped genes reside in previously established murine homeobox gene clusters. Specifically, two map to the cluster HOX-1 on chromosome 6 and one each to HOX-3 and HOX-4 on chromosome 15 and 2, respectively. The ratio of newly identified homeobox genes to the previously characterized murine homeobox genes suggests that there remain several uncharacterized homeobox genes in the murine genome.**

Molecular and genetic analyses in *Drosophila* have identified sets of genes that are crucial in pattern formation (reviewed in ref. 1). A distinguishing feature of one group of these genes is the presence of a 180-base-pair region, the homeobox, that encodes the 60-amino acid homeodomain (2). There is increasing evidence from transcription and transfection studies that homeobox genes can trans-regulate their own transcription as well as that of other homeobox genes via homeodomain-mediated sequence-specific DNA binding (3, 4). Moreover, it is believed that homeobox genes control the temporal and spatial expression patterns of arrays of yet unidentified target genes (1, 3, 4).

Interestingly, the homeodomain amino acid residues are conserved across species including nematodes, annelids, arthropods, vertebrates, and plants (1, 5–7), implying functional constraints for their retention (1, 8). The high degree of conservation has aided in identifying and isolating homologues of *Drosophila* homeobox genes from different phylogenetic species by cross-hybridizing *Drosophila* homeobox sequences.

Over 30 murine homeobox (Hox) genes have been identified, with most of these genes organized in four clusters on four different chromosomes (reviewed in refs. 9 and 10). Presumably the four HOX clusters arose by duplication of an ancestral cluster during the evolution of the vertebrate lineages (11, 12). In addition to the four clusters of Hox genes, other somewhat divergent Hox genes including *Hox-7* (13, 14), *Cdx* (15), *Evx* (16), and the *En* (17) types have been found individually dispersed on different chromosomes of the murine genome. Most of the Hox genes have been isolated by hybridization with other Hox gene probes and/or by isolation of overlapping genomic clones. However, because of the degeneracy of the genetic code, even homeobox genes specifying very similar amino acid sequences may fail to cross-hybridize. For example, early Southern hybridization experiments failed to detect the presence of any homeobox genes in the genome of nematodes (8, 18), although a subsequent screening with a degenerate oligonucleotide probe revealed the presence of about 60 such homeobox genes (5). To identify previously uncharacterized murine Hox genes we screened a mouse genomic library with a degenerate oligomer corresponding to a highly conserved motif in the homeodomain. Our screening has led to the sequencing of the homeodomain region of ten novel Hox genes and suggests that there are many yet unidentified Hox genes in the murine genome.§

MATERIALS AND METHODS

Genomic Screening. A 768-fold degenerate 23-mer oligonucleotide (H23) was used to screen a commercially available adult mouse DBA/2J genomic DNA library in λ phage EMBL-3 (Clontech) for murine Hox genes. About 3×10^5 plaques representing approximately one haploid murine genome equivalent sequences were plated. To increase the signal-to-noise ratio, the plaques were amplified on bacteria-impregnated nitrocellulose filters (19), processed, and hybridized with labeled H23 oligonucleotide (specific activity, 10^9 cpm/μg) for 60–80 hr at 42°C. The filters were washed in 3 M tetramethylammonium chloride/50 mM Tris/0.2 mM EDTA, pH 8.0, at ambient temperature, 37°C, and 56°C for 30 min each (20). DNA isolation, restriction enzyme digestion, agarose gel electrophoresis, dot blot, Southern blot transfer, and dideoxy sequencing were performed as described (21).

Chromosomal Mapping, Probes, and Restriction Fragment Length Polymorphisms (RFLPs). Interspecific backcross progeny were generated by mating (C57BL/6J × *Mus spretus*) F_1 females and C57BL/6J males. A total of 205 N_2 progeny were obtained; a random subset of these N_2 mice were used to map each of the new Hox loci. The Hox probes and RFLPs used for mapping were as follows. For the Hox clones 49h, 61h, 13h, 20h, 23h, 9t, 15h, and 71h the respective probe segments 2.0-kilobase (kb) *Hin*dIII, 1.1-kb *Eco*RI–*Hin*dIII, 1.7-kb *Eco*RI, 1.1-kb *Eco*RI, 2.6-kb *Eco*RI, 1.8-kb *Eco*RI–*Hin*dIII, 1.1-kb *Bam*HI–*Xba* I and 3.3-kb *Bam*HI detected hybridizing fragments of 11.0, 6.5, 7.0, 2.0 and 1.2,

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

Abbreviations: Hox, murine homeobox; RFLP, restriction fragment length polymorphism; TMAC, tetramethylammonium chloride.

†To whom reprint requests should be addressed.

§The sequences reported in this paper have been deposited in the GenBank data base (accession nos. A38809 to I38809 and A38810).

7.8 and 3.2, 9.0 and 7.0, 3.3 and 0.7, and 14 kb in C57BL/6J genomic DNA and segments of 5.8, 4.5, 14.0, 13.1, 8.8, 5.3, 5.6, and 11.0 kb in *M. spretus* DNA digested with *Xba* I, *Pst* I, *Bgl* I, *Bgl* II, *Bam*HI, *Sac* I, *Pst* I, and *Xba* I, respectively.

The chromosomal positions of several loci used herein have not yet been reported for our interspecific backcross. These include endothelial leukocyte adhesion molecule 1 (*Elam*) on chromosome 1, multidrug resistance 1 (*Mdr-1*) and erythropoietin (*Epo*) on chromosome 5, and *Hox-3.1* on chromosome 15. The probes and RFLPs used to map these loci were as follows. The *Elam* probe, a 3.5-kb human cDNA (22), identified fragments of 2.9, 2.1, and 1.4 kb (B6) and 2.5, 1.6, and 1.4 kb (S) in *Taq* I-digested DNA. The *Mdr-1* probe, a 4.7-kb hamster cDNA (23), identified fragments of 22.0 kb (B6) and 8.0 kb (S) in *Bgl* I-digested DNA. The *Epo* probe, a 1.0-kb mouse cDNA (24), identified fragments of 8.2 kb (B6) and 6.6 kb (S) in *Xba* I-digested DNA. The *Hox-3.1* probe, a 350-base-pair (bp) segment (25), identified fragments of 1.0 and 0.5 kb (B6) and 1.4 and 0.2 kb (S) in *Msp* I-digested DNA. Finally, the *Pdgfa* probe, a 906-bp mouse cDNA segment, identified fragments of 3.4, 2.4, 1.9, 1.2, and 0.5 kb (B6) and 4.8, 2.8, 1.9, and 0.5 kb (S) in *Pvu* II-digested DNA. This locus mapped proximal to *Epo* on chromosome 5.

Recombinant distances were calculated (26) using the computer program SPRETUS MADNESS developed by D. Dave and A. M. Buchberg (National Cancer Institute, Frederick, MD). Gene order was determined by minimizing the number of recombinant events required to explain the allele distribution patterns.

RESULTS AND DISCUSSION

Optimizing Washing Conditions to Select Hox Genes. A 768-fold degenerate 23-base oligonucleotide (H23) corresponding to a highly conserved 8-amino acid motif in the recognition helix of most homeodomains was synthesized (Fig. 1*A*). H23 and its reverse complement detected only Hox genes at 0–3 mismatches when the Microgenie data bank (Beckman) was searched. With 4 mismatches, however, exon three of a non-homeobox gene, the murine thymidylate synthase gene (27), was selected.

To optimize hybridization and washing conditions for preferentially selecting Hox genes over non-Hox genes, we

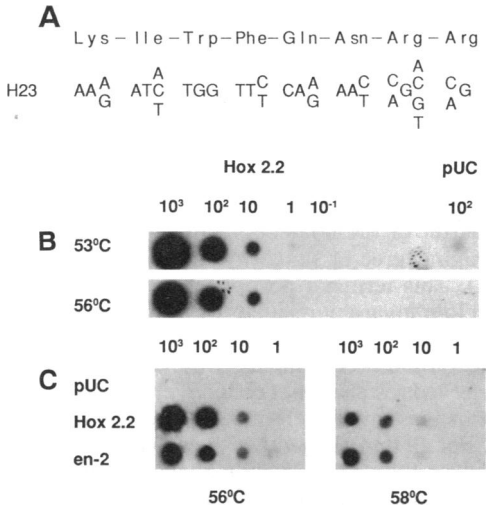

FIG. 1. (*A*) Sequence of the 23-base, 768-fold degenerate oligonucleotide (H23) corresponding to the highly conserved 8-amino acid motif of homeodomain proteins. (*B* and *C*) Various amounts (10^3 to 1 or 10^{-1} ng) of *Hox-2.2* (p611), *En-2* (p12rt), and pUC linearized DNAs in 10 μl were spotted on nitrocellulose membrane, processed, and hybridized to H23 oligonucleotide at 37°C. Washing was in 3 M TMAC at 53°C, 56°C, or 58°C.

equivalently dot-blotted various concentrations of linearized p611, a recombinant plasmid containing the homeobox region of the *Hox-2.2* gene (0 mismatch) and pUC, a non-homeobox-containing plasmid as a negative control (Fig. 1*B*). The pUC DNA contains a region of H23 homology with 5 mismatches. The dot blots were processed, hybridized to H23, and then washed in 3 M TMAC to allow an empirical determination of the duplex melting temperature on the basis of unit length as opposed to base composition (20). Test washes were at 53°C, 56°C, and 58°C. When washed at 53°C, 100 ng of pUC DNA retained a hybridization signal, albeit about 50 times less intense in comparison to the signal from an equivalent amount of *Hox-2.2* DNA (Fig. 1*B*). However, washing in 3 M TMAC at 56°C completely eliminated this nonspecific hybridization but still allowed selection of Hox genes with a single mismatch as demonstrated by hybridization of p12rt, a recombinant plasmid containing the homeobox region of the murine *En-2* gene (Fig. 1*C*). Since more stringent washing (58°C) resulted in a weaker signal, we chose to wash the blots at 56°C.

Identification of Murine Hox Genes. We screened approximately one haploid genome equivalent of a murine genomic library with H23. This degenerate oligomer would detect known murine homologs of *Antp, cad, Dll, Dfd, lab, ftz, msh, ro, scr*, and *zen* types of *Drosophila* Hox genes at 0 mismatch and *en, eve, bcd*, and *H2.0* types at a single mismatch. However, it would not be expected to detect diverse Hox genes, including murine homologs of *ct* (up to 5 mismatches), *prd* (up to 6 mismatches), and POU (up to 8 mismatches) types or Hox genes that are interrupted by an intron in the 8-amino acid motif.

We isolated 70 positive clones that were twice plaque-purified. To facilitate Phagescript (Stratagene) subcloning and sequencing, approximately 0.2- to 1.0-kb *Hae* III, *Hin*fI, and *Taq* I DNA fragments that hybridized with the H23 oligonucleotide were identified. A total of 29 clones were randomly selected for sequencing. The sequence analysis indicated that all the clones carried 0–2 mismatches with the degenerate oligonucleotide probe. Twelve of the 29 sequenced clones proved to be duplicates, with the remaining 17 distinct clones falling into three broad categories. Six clones represented previously characterized murine Hox genes. These clones contained nucleotide sequences identical to *En-2* and *Hox-2.1, -2.4, -3.1, -3.2*, and *-4.1* genes, thus confirming the effectiveness of the degenerate oligonucleotide hybridization/washing conditions in identifying genuine Hox genes. In addition, 10 of the clones sequenced appeared to be novel murine Hox genes and a single clone an artifact.

Sequence analysis indicated that all the clones representing novel Hox genes encoded amino acid sequences with considerable homology to the homeodomain consensus sequence. This was particularly evident in the 7 clones that contained sequences corresponding to the entire homeodomain (Fig. 2*A*). All of these clones showed identity to the consensus sequence at a minimum of 19 of 21 amino acid positions that are known to be highly conserved. And at other conserved positions, conservative changes were observed. Moreover, another large conserved block, of 7 amino acids, (L/R)ELE(K/R)EF, present in helix 1 occurred in all the full-length homeodomains at positions 14–20. These observations strongly suggest that these seven clones represent true Hox genes. Based upon sequence similarity these homeodomains could be further arranged into four distinct groups (Fig. 2*A*).

For three clones (15h, 71h, and 60h) there was no significant amino acid sequence homology to the consensus homeodomain 5′ of the region complementary to H23. Nevertheless, analysis of the data indicates that these are also true Hox genes (Fig. 2*A*). These three loci carry in-frame splice acceptor consensus sequences at the junctions where homol-

A

```
                          Helix 1          Helix 2      Helix 3 Helix 4
                         ----------       ----------   ----------------
                                                        ** * *
C.S.:          ----R--Y---Q---L---F----Y-----R---A--L-L---Q-KIWFQNRR-K-K---
                              ▲        ▲        ▲        ▲        ▲        ▲
                         10        20        30        40        50        60

49h:  Hox-4.3    RRRG.QT.SRF.TLE.EKE.LFNP.LTRKR.IEVSHT.A.TER.V........M.W.KEN

                                                      +++o
 55t:  Hox-3.7   GRKK.CP.TKH.TLE.EKE.LFNM.LTRER.LEISKTIN.TDR.V........M.L.KMN
 61h:  Hox-1.8   GRKK.CP.TKH.TLE.EKE.LFNM.LTRER.LEISRSVH.TDR.V........M.L.KMN

                      + +          o
 13h:  Hox-3.8   TRKK.CP.SKF.IRE.ERE.FFNV.INKEK.LQLSRM.N.TDR.V........M.E.KLS
 20h:  Hox-1.9   TRKK.CP.TKY.IRE.ERE.FFSV.INKEK.LQLSRM.N.TDR.V........M.E.KIN

                 o                 o
 23h:  Gsh-1     SKRM.TAFTST.LLE.ERE.ASNM.LSRLR.IEI.TY.N.SEK.V........V.H.KEG
  9t:  Gsh-2     GKRM.TAFTST.LLE.ERE.SSNM.LSRLR.IEI.TY.N.SEK.V........V.H.KEG
 15h:  Gsh-3                                                <V........S.F.KMW
 71h:  Gsh-4                                                  <V.......A.E.RLK
 60h:  Gsh-5                                                <V........M.WRNSK

 39h:            YQYGVUWELTRRNREVWVFLTCVKUMWTNYLYEILNLLNSITNYL.......TNGAFYCL
```

B

```
                                                                      +
13h:  Hox-3.8   TRKK.CP.SKF.IRE.ERE.FFNV.INKEK.LQLSRM.N.TDR.V........M.E.KLS
 3H:            TRKK.CP.SKF.IRE.ERE.FFNV.INKEK.LQLSRM.N.TDR.V........M.D.KLS

20h:  Hox-1.9   TRKK.CP.TKY.IRE.ERE.FFSV.INKEK.LQLSRM.N.TDR.V........M.E.KIN
 1I:            TRKK.CP.TKY.IRE.ERE.FFSV.INKEK.LQLSRM.N.TDR.V........M.E.KIN

                                                  o
55t:  Hox-3.7   GRKK.CP.TKH.TLE.EKE.LFNM.LTRER.LEISKTIN.TDR.V........M.L.KMN
 3I:            GRKK.CP.TKH.TLE.EKE.LFNM.LTRER.LQISKTIN.TDR.V........M.L.KMN

61h:  Hox-1.8   GRKK.CP.TKH.TLE.EKE.LFNM.LTRER.LEISRSVH.TDR.V........M.L.KMN
 1H:            GRKK.CP.TKH.TLE.EKE.LFNM.LTRER.LEISRSVH.TDR.V........M.L.KMN

                                                  o
49h:  Hox-4.3   RRRG.QT.SRF.TLE.EKE.LFNP.LTRKR.IEVSHT.A.TER.V........M.W.KEN
 4E:            RRRG.QT.SRF.TLE.EKE.LFNP.LTRKR.IEVSHA.A.TER.V........M.W.KEN

                          oo                      o
49h:  Hox-4.3   RRRG.QT.SRF.TLE.EKE.LFNP.LTRKR.IEVSHT.A.TER.V........M.W.KEN
Hox-4.3         RRRG.QT.SRF.RVE.EKE.LFNP.LTRKR.IEVSHS.A.TER.V........M.W.KEN
Hox-2.4         RRRG.QT.SRY.TLE.EKE.LFNP.LTRKR.IEVSHA.G.TER.V........M.W.KEN
Hox-3.1         RRSG.QT.SRY.TLE.EKE.LFNP.LTRKR.IEVSHA.G.TER.V........M.W.KEN
```

C

	Hox-1.4 (33)	Hox-1.5 (34)	Hox-2.4 (32)	Hox-1.7 (35)	En-1 (17)	Hox-7 (13,14)	Evx-1 (16)	Cdx-1 (15)	Hox-1.9
23h (Gsh-1) 9t (Gsh-2)	40[46]	38[46]	34[42]	35[43]	32[36]	31[37]	31[39]	28[38]	27[39]

FIG. 2. (*A*) Alignment of deduced homeodomain sequences (amino acids in single-letter code) of 11 clones with the consensus sequence (C.S.). The consensus sequence was obtained by comparative analysis of 83 different homeodomain sequences (2). Invariant residues (stars), conserved residues, and the positions of variable residues (dashes) are shown. The positions of four α-helices identified by NMR spectroscopy (28) are shown at the top. Dots represent amino acids identical to the conserved residues in the consensus sequence. The positions of introns in clones 15h, 60h, and 71h are indicated (<). Related homeodomains are paired together. Also, six previously characterized Hox genes (*En-2* and *Hox-2.1, -2.4, -3.1, -3.2,* and *4.1*) were identified in the screening. Clone 39h, selected at 2-base mismatch, appeared to be an artifact. Chemically similar amino acids (S and T; I, V, and L; F and Y; and R and K) among the pairs are indicated by a plus sign and dissimilar exchanges are shown by open circles. (*B*) Comparison of homeodomain sequences of the clones to the related murine and human homeodomain sequences. Homeodomain sequence of murine paralogs 13h (*Hox-3.8*) and 20h (*Hox-1.9*) are compared with their human cognate genes (29, 30); 55t (*Hox-3.7*) and its paralog 61h (*Hox-1.8*) are compared with their respective human cognate sequences (29, 30); 49h (*Hox-4.3*) is compared with independently described *Hox-4.3* (31), with its paralogs, *Hox-2.4* (32) and *Hox-3.1* (25), and with the human cognate, *4E* (29, 30). Chemically similar amino acids (S and T; D and E) among the heterologous cognate pairs are indicated by a plus sign and dissimilar exchanges are shown by open circles. (*C*) Sequence comparison of 23h (*Gsh-1*) and 9t (*Gsh-2*) homeodomains with other murine homeodomains. The total number of amino acid residues that are identical or chemically similar (shown in brackets) in pairwise comparisons are indicated. References are in parentheses.

ogy is lost, suggesting the presence of introns within these Hox sequences. The splice acceptor sequence observed for 60h (NTAG/G) is relatively more common than that found in 15h (NAAG/G) (36). Although final confirmation of these splice sites must await comparison of the genomic sequences with their respective cDNAs, nevertheless it is interesting that the previously characterized homeodomains of the *Drosophila lab* (37), *Dll* (38), *Abd B* (39), and *NK-1* (40) genes and *Caenorhabditis elegans ceh-2* and *ceh-7* genes (5) are interrupted by an intron prior to the same position-46 valine as clones 60h and 15h. The remaining Hox locus (71h) was least homologous to the degenerate oligonucleotide probe. Selected at 2-base mismatch, this clone had only 6 residues in common with the 8-amino acid motif (Fig. 2*A*). Apparently, the nucleotide sequence corresponding to the first 2 residues encompasses the splice acceptor site (NCAG/G). Although an intron disrupts the 8-amino acid motif, the locus was probably selected due to a fortuitous match between intronic sequences and H23. Based on the acceptor site, the intron of 71h was present before the valine-48 codon, as is noticed in the homeodomains of *C. elegans ceh-4, ceh-8,* and *ceh-14* (5); murine *Evx-1* and -2 (15), and their *Xenopus* homologue, *XHox3* (41). The presence of introns in these three loci permitted sequence comparison to only the carboxyl quarter of the homeodomain, which nonetheless did show a high degree of similarity to the consensus sequence (Fig. 2*A*). This terminal portion of the homeodomain contained all four invariant residues (tryptophan, phenylalanine, asparagine, and arginine at positions 48, 49, 51, and 53, respectively) and several highly conserved residues as well as less conserved residues that collectively form the two tightly linked α-helices implicated in sequence-specific DNA binding (28). For all three loci either the conserved residues were preserved or their conservative replacements were found, within and flanking the 8-amino acid motif. For example in 71h, a valine was present instead of isoleucine at position 47 and in 60h, an arginine replaced lysine at position 57. Moreover, most of the amino acid residues present at the variable positions have been noted in other homeodomains. Collectively these observations suggest that these loci represent bona fide Hox sequences. The analyzed homeodomain regions for Hox loci 15h, 60h, and 71h are nevertheless too

small (<16 amino acids) for us to make any conclusive statements about their relatedness to any of the previously established groups. Finally, only a single clone (39h) out of the 29 clones sequenced was not a Hox gene, as it lacked significant open reading frame, did not encode one of the invariant amino acids, and did not show homology to the homeodomain consensus sequence outside of the 8-amino acid motif. The high percentage (97%) of genuine Hox genes selected suggests that degenerate oligomer hybridization may be the method of choice in identifying members of multigene families from complex mammalian genomes.

Comparison of the seven full-length homeodomains described above with the various types of previously characterized homeodomains suggests that 23h and 9t represent a new subfamily of Hox genes (Fig. 2*A*). The homeodomain sequences of 23h and 9t are closely related to each other (2 mismatches), but show 20–33 mismatches in pairwise comparisons with other, previously sequenced homeodomains (Fig. 2*C*). Nevertheless, it is interesting that the 23h and 9t homeodomains are approximately as similar to the *Antp* homeodomain of *Drosophila*, with 37 of 60 amino acids identical, as the new HOX cluster-associated Hox genes reported in this paper, which have 29–39 of 60 amino acids identical.

On the other hand, the homeodomain sequences of 61h and 20h were completely identical to those of two adjacently placed human homeobox genes, *1H* and *1I*, respectively (29) (Fig. 2*B*). Furthermore, the homeodomains of 13h and 55t were highly homologous (98% amino acid identity) to those of two other adjacently placed human homeobox genes, *3H* and *3I*, respectively (30). This indicated that these newly identified murine Hox genes are cognates of previously described human homeobox genes. Because of the high degree of similarity between the homeodomains of murine and human Hox genes and the congruence in the overall structural organization of murine and human Hox gene clusters, one could tentatively assign the newly identified Hox sequences to equivalent positions on the murine HOX clusters, placing 61h and 20h within HOX-1 and placing 55t and 13h within HOX-3 (Fig. 3*A*).

Hox clone 49h showed significant homology (98%) to human homeobox gene *4E* (29) (Fig. 2*B*) allowing tentative assignment of this locus to the *Hox-4.3* position, paralogous to *Hox-2.4* (32) and *Hox-3.1* (25). As expected between paralogous genes, the 49h homeodomain is highly homologous to murine *Hox-2.4* (95%) and *Hox-3.1* (93%). Surprisingly, the 49h homeodomain showed higher homology to its human cognate *4E* (29) than to the *Hox-4.3* cDNA sequence published in 1990 (31). The sequence presented herein more closely matches the consensus for this paralogous group at two highly conserved amino acid positions (Fig. 2*B*), suggesting that the observed discrepancies are not merely the result of strain differences.

Chromosomal Mapping of the Hox Genes. The cluster assignments of all these clones, except 55t and 60h, have been confirmed by chromosome mapping. Single-copy sequences from these clones were hybridized to an interspecific backcross panel derived by crossing C57BL/6J and *M. spretus*. This panel has been typed for over 675 loci that are well distributed among all the autosomes as well as the X chromosome (45). C57BL/6J and *M. spretus* DNAs were digested with several restriction endonucleases and analyzed by Southern blot hybridization for informative RFLPs with each of the Hox probes. The strain distribution pattern of each RFLP in the interspecific backcross mice was then determined and used to position the Hox genes on the interspecific map (Fig. 3).

Four of the Hox loci cosegregated with known Hox genes, consistent with cluster assignments based on sequence similarities to human Hox genes. These loci included clone 49h,

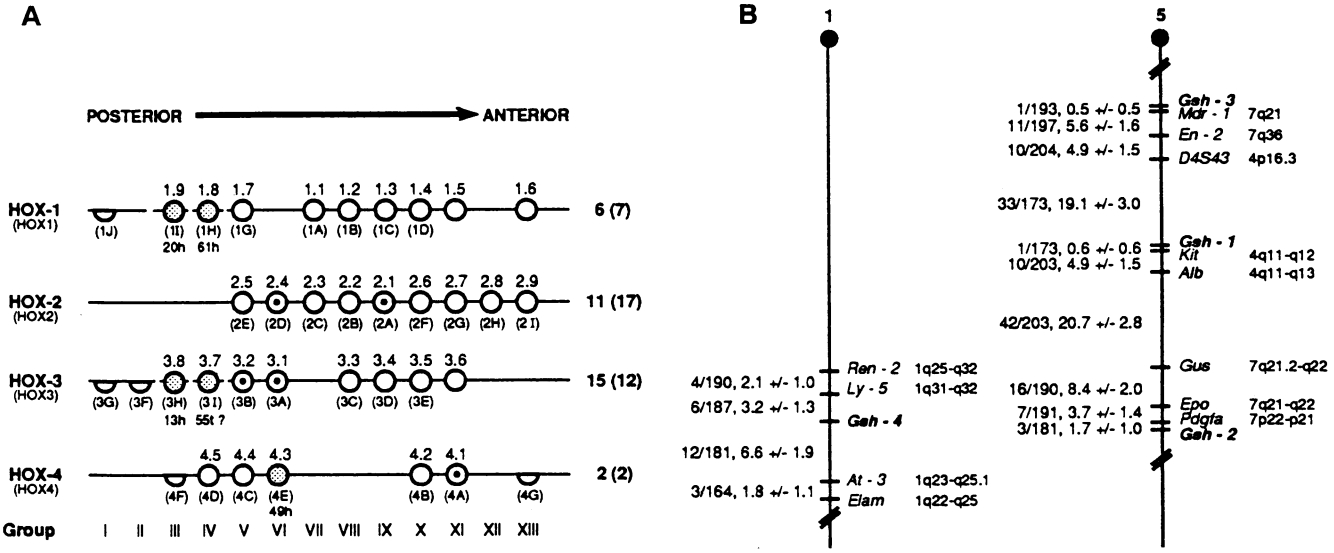

FIG. 3. (*A*) Structural organization of the four murine and human (in parentheses) HOX clusters (not drawn to scale). Murine and human Hox genes are designated according to the proposed nomenclature (42, 43). The genes on the HOX clusters are vertically aligned to highlight 13 groups (I–XIII) of paralogous genes. Stippled circles indicate the murine Hox genes characterized in this study. Both 20h and 61h map to the HOX-1 cluster, whereas single-copy sequences from 13h and 49h map to HOX-3 and HOX-4, respectively. Clone 55t is tentatively assigned to the HOX-3 cluster on the basis of extensive sequence similarity to its human cognate, *3I* (30). Circles with a dot represent previously known murine Hox genes identified in our screening, open circles indicate other sequenced murine and human Hox genes, and semicircles indicate positions of human Hox genes for which a corresponding mouse gene has not yet been described (29, 30). Human cognates of murine *Hox-1.5*, *-1.6*, and *-3.6* are yet to be identified. (*B*) Linkage maps showing the chromosomal locations of four Hox loci (*Gsh-1* to *-4*) that did not cosegregate with any of the known murine Hox genes. The Hox loci were mapped by interspecific backcross analysis. The number of recombinant N_2 animals typed plus the recombinant distance in centimorgans (± 1 standard error) is shown for each pair of loci on the left of the chromosome maps. Where no recombinants were found between loci, the 95% confidence limit of the recombinant distance is given. The positions of loci on human chromosomes are shown to the right of the chromosome maps. The *Ren-2*, *Ly-5*, and *At-3* loci have been positioned on chromosome 1 (44) whereas *En-2*, *D4S43* (an anonymous DNA marker), *Kit*, *Alb*, and *Gus* have been localized on chromosome 5 (13).

which cosegregated with *Hox-4.2* on chromosome 2; clones 20h and 61h, which cosegregated with *Hox-1.3* on chromosome 6; and clone 13h, which cosegregated with *Hox-3.1* on chromosome 15. Based on the chromosomal mapping and sequence similarity to the preassigned human Hox genes, the murine Hox loci 61h and 20h were designated *Hox-1.8* and *Hox-1.9* respectively; similarly, Hox loci 13h and 49h were designated *Hox-3.8* and *Hox-4.3*, respectively, and Hox locus 55t was tentatively designated *Hox-3.7*. The remaining five Hox clones, 9t, 23h, 15h, 71h and 61h, were designated *Gsh-1* to *Gsh-5* (*g*enomic *s*creening *h*omeobox), respectively.

Chromosomal mapping of *Gsh-1* to *Gsh-4* indicated that none of these loci cosegregated with any of the previously known Hox genes, again consistent with sequence comparison data. The chromosomal positions of these four dispersed Hox loci are shown in Fig. 3*B*. Alignment of the interspecific map with the composite map provided by GBASE (The Jackson Laboratory) failed to identify any known mouse mutations that map in the vicinity of the Hox loci with a phenotype consistent with what might be expected for a defect in a Hox gene.

The cloning of only a fraction (about 20%) of the previously identified Hox genes indicates that only a subset of total Hox genes that are recognizable by the degenerate probe have been characterized in this screen. However, it is noteworthy that the ratio of new to previously characterized Hox genes identified with the degenerate probe is about 1.7, and at least four of these novel Hox genes reside at chromosomal positions other than the four main clusters of murine Hox genes. If this degenerate oligonucleotide screen has indeed selected a random subset of all Hox genes present in the genome, then at least 15 yet-undiscovered Hox genes that are complementary to our probe remain to be isolated.

G.S. and S.K. made equivalent contributions to this work. The chromosomal mapping probes were kindly provided by K. McLain (*Hox-1.9*), M. Bevilacqua (*Elam*), I. B. Roninson (*Mdr*-1), E. Goldwasser (*Epo*), B. Mock (*Hox-3.1*), and U. Drager (*Pdgfa*). We thank them and also extend our appreciation to T. Valerius, P. Potter, D. Swing, C. Silan, M. Adams, D. Angle, and B. Cho for excellent technical assistance. The help of V. Pancholi and M. Kern in acquisition of GenBank sequences is gratefully acknowledged. This research was supported in part by National Cancer Institute Contract NO1-CO-74101 with Advanced BioScience Laboratories (N.A.J. and N.G.C.) and National Institutes of Health Grant HD24517 (S.S.P.). S.K. is a Ryan fellow and G.S. is a Children's Hospital Medical Center Board of Trustees fellow.

1. Scott, M. P., Tamkun, J. W. & Hartzell, G. W., III (1989) *Biochim. Biophys. Acta* **989**, 25–48.
2. Scott, M. P. & Weiner, A. J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 4115–4119.
3. Hoey, T. & Levine, M. (1988) *Nature (London)* **332**, 858–861.
4. Winslow, G. M., Hayashi, S., Krasnow, M., Hogness, D. S. & Scott, M. P. (1989) *Cell* **57**, 1017–1030.
5. Bürglin, T. R., Finney, M., Coulson, A. & Ruvkun, G. (1989) *Nature (London)* **341**, 239–243.
6. Weeden, C. J., Kostriken, R. G., Matsumura, I. & Weisblat, D. A. (1990) *Nucleic Acids Res.* **18**, 1908.
7. Vollbrecht, E., Veit, B., Sinha, N. & Hake, S. (1991) *Nature (London)* **350**, 241–243.
8. Holland, P. W. H. & Hogan, B. L. M. (1986) *Nature (London)* **321**, 251–253.
9. Kessel, M. & Gruss, P. (1990) *Science* **249**, 374–379.
10. Graham, A., Papalopulu, N. & Krumlauf, R. (1989) *Cell* **57**, 367–378.
11. Kappen, C., Schughart, K. & Ruddle, F. H. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5459–5463.
12. Schughart, K., Kappen, C. & Ruddle, F. H. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7067–7071.
13. Hill, R. E., Jones, P. F., Rees, A. R., Christina, M. S., Justice, M. J., Copeland, N. G., Jenkins, N. A., Graham, E. & Davidson, D. R. (1989) *Genes Dev.* **3**, 26–37.
14. Robert, B., Sassoon, D., Jacq, B., Gehring, W. & Buckingham, M. (1989) *EMBO J.* **8**, 91–100.
15. Duprey, P., Chowdhury, K., Dressler, G. R., Balling, R., Simon, D., Guenet, J. & Gruss, P. (1988) *Genes Dev.* **2**, 1647–1654.
16. Bastian, H. & Gruss, P. (1990) *EMBO J.* **9**, 1839–1852.
17. Joyner, A. L. & Martin, G. R. (1987) *Genes Dev.* **1**, 29–38.
18. McGinnis, W. (1985) *Cold Spring Harbor Symp. Quant. Biol.* **50**, 263–270.
19. Woo, S. L. C. (1979) *Methods Enzymol.* **68**, 389–395.
20. Wood, W. I., Gitschier, J., Lasky, L. & Lawn, R. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1585–1588.
21. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
22. Bevilacqua, M. P., Stengelin, S., Gimbrone, M. A. & Seed, B. (1989) *Science* **243**, 1160–1165.
23. Roninson, I. B., Chin, J. E., Choi, K., Gros, P., Housman, D. E., Fojo, A., Shen, D. W., Gottesman, M. M. & Pastan, I. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 4538–4542.
24. McDonald, J. D., Lin, F. K. & Goldwasser, E. (1986) *Mol. Cell Biol.* **6**, 842–848.
25. Breier, G., Dressler, G. R. & Gruss, P. (1988) *EMBO J.* **7**, 1329–1336.
26. Green, E. L. (1981) in *Genetics and Probability in Animal Breeding Experiments* (Macmillan, New York), pp. 77–113.
27. Deng, T., Li, D., Jenh, C. & Johnson, L. F. (1986) *J. Biol. Chem.* **261**, 16000–16005.
28. Otting, G., Qian, Y., Billeter, M., Muller, M., Affolter, M., Gehring, W. & Wüthrich, K. (1990) *EMBO J.* **9**, 3085–3092.
29. 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**, 10385–10403.
30. Stornaiuolo, A., Acampora, D., Pannese, M., D'Esposito, M., Morelli, F., Migliaccio, E., Rambaldi, M., Faiella, A., Nigro, V., Simeone, A. & Boncinelli, E. (1990) *Cell Differ. Dev.* **31**, 119–127.
31. Izpisua-Belmonte, J., Dollé, P., Renucci, A., Zappavigna, V., Falkenstein, H. & Duboule, D. (1990) *Development* **110**, 733–745.
32. Hart, C. P., Fainsod, A. & Ruddle, F. H. (1987) *Genomics* **1**, 182–195.
33. Rubin, M. R., Toth, L. E., Patel, M. D., D'Eustachio, P. & Nguyen-Huu, M. C. (1986) *Science* **233**, 663–667.
34. McGinnis, W., Hart, C. P., Gehring, W. J. & Ruddle, F. H. (1984) *Cell* **38**, 675–680.
35. Rubin, M. R., King, W., Toth, L. E., Sawczuk, I. S., Levine, M. S., D'Eustachio, P. & Nguyen-Huu, M. C. (1988) *Mol. Cell Biol.* **8**, 5593.
36. Mount, S. (1982) *Nucleic Acids Res.* **10**, 459–474.
37. Mlodzik, M., Fjose, A. & Gehring, W. J. (1988) *EMBO J.* **7**, 2569–2578.
38. Cohen, S. M., Brönner, G., Küttner, F., Jürgens, G. & Jackle, H. (1989) *Nature (London)* **338**, 432–434.
39. Celniker, S. E., Keelan, D. J. & Lewis, E. B. (1989) *Genes Dev.* **3**, 1424–1436.
40. Kim, Y. & Nirenberg, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7716–7720.
41. Ruiz i Altaba, A. & Melton, D. A. (1989) *Development* **106**, 173–183.
42. Duboule, D., Boncinelli, E., DeRobertis, E., Featherstone, M., Lonai, P., Guillermo, O. & Ruddle, F. H. (1990) *Genomics* **7**, 458–459.
43. McAlpine, P. J. & Shows, T. B. (1990) *Genomics* **7**, 460.
44. Dickinson, M. E., Kobrin, M. S., Silan, C. M., Kingsley, D. M., Justice, M. J., Miller, D. A., Ceci, J. D., Lock, L. F., Lee, A., Buchberg, A. M., Siracusa, L. D., Lyons, K. M., Derynck, R., Hogan, B. L. M., Copeland, N. G. & Jenkins, N. A. (1990) *Genomics* **6**, 505–520.
45. Copeland, N. G. & Jenkins, N. A. (1991) *Trends Genet.* **7**, 113–118.