Volume 17 Number 24 1989 Nucleic Acids Research

The human HOX gene family

Dario Acampora, Maurizio D'Esposito, Antonio Faiella, Maria Pannese, Enrica Migliaccio, Franco Morelli, Anna Stornaiuolo, Vincenzo Nigro, Antonio Simeone and Edoardo Boncinelli*

International Institute of Genetics and Biophysics, CNR, Via Marconi 10, 80125 Naples, Italy

Received September 27, 1989; Revised and Accepted November 16, 1989 EMBL accession nos X16665–X16667

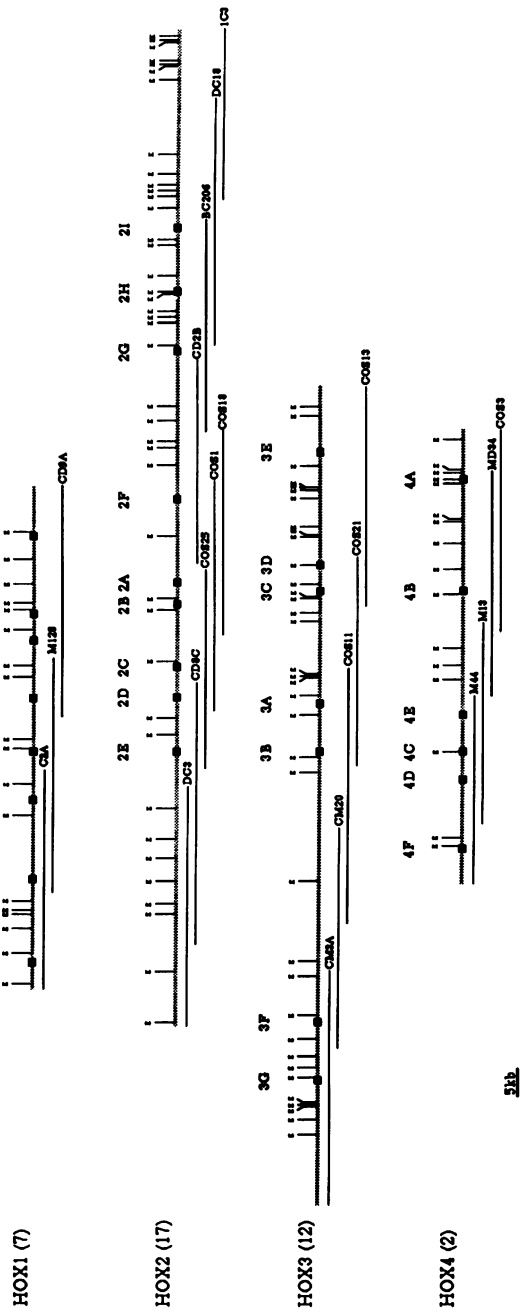
Abstract

We report the identification of 10 new human homeobox sequences. Altogether, we have isolated and sequenced 30 human homeoboxes clustered in 4 chromosomal regions called HOX loci. HOX1 includes 8 homeoboxes in 90 kb of DNA on chromosome 7. HOX2 includes 9 homeoboxes in 180 kb on chromosome 17. HOX3 contains at least 7 homeoboxes in 160 kb on chromosome 12. Finally, HOX4 includes 6 homeoboxes in 70 kb on chromosome 2. Homeodomains obtained from the conceptual translation of the isolated homeoboxes can be attributed to 13 homology groups on the basis of their primary peptide sequence. Moreover, it is possible to align the 4 HOX loci so that corresponding homeodomains in all loci share the maximal sequence identity. The complex of these observations supports and extends an evolutionary hypothesis concerning the origin of mammalian and fly homeobox gene complexes. We also determined the coding region present in 3 HOX2 cDNA clones corresponding to HOX2G, HOX2H and HOX2I.

Introduction

Genes containing homeobox sequences encode nuclear proteins with regulatory functions in a wide variety of organisms, from yeast to man (1,2). Many of these genes function in the control of early developmental programs in Drosophila and most likely in other organisms. Evidence is accumulating that the encoded homeoproteins function as transcription factors regulating the expression of other genes and homeodomain binding sites have been identified in the cis-regulatory sequences of genes known to be regulated by homeobox genes (3,4). On the other hand, homeoproteins have been shown to be transcriptional activators of the growth hormone and prolactin genes in the pituitary and to bind to the octanucleotide sequence present in many mammalian promoters and enhancers (5,6).

Although the DNA-binding domain specified by the homeobox, termed homeodomain (HD), is clearly an ancient and conserved functional motif, homeobox gene families have evolved encoding HDs with different primary sequences. Indeed, known HDs can be attributed to different classes according to their primary sequence. Class I HDs are the most closely related to the


HOX1 (7)
HOX2 (17)
HOX3 (12)
HOX4 (2)
5kb

Antennapedia (Antp) HD and have also been designated Antp-like HDs (7). In Drosophila, genes containing a class I HD are clustered in two complex loci, the Antennapedia-complex (ANT-C) and the bithorax-complex (BX-C) (8). Mouse and human class I homeobox genes appear to be clustered in a similar way in restricted genomic regions (HOX loci) of at least four chromosomes (9,2).

We previously reported the isolation of 20 human homeoboxes (9). We report here the identification of 10 new homeoboxes and summarize the genomic organization of 30 human homeoboxes so far identified in the four HOX1, 2, 3 and 4 loci. These HOX loci appear to be homologous to each other and, to a more limited extent, to the Drosophila homeotic gene complexes.

Materials and Methods

A human genomic library in pcos2EMBL cosmid vector (10) was kindly provided by Anna-Maria Frischauf. Two cDNA libraries in λgt11 (11) were prepared from N-TERA2 cells derived from a human teratocarcinoma treated with 10 μM retinoic acid for 14 days (12,13). All libraries were screened according to standard procedures (14). We previously published partial genomic maps (16,17,9) of the four HOX loci. These were obtained by chromosome walking around the homeobox sequences first isolated, i.e. HOX1D, 2C, 3C and 4B. We used pcos2EMBL cosmid clones throughout and did not find any inconsistency in comparing maps derived from cosmid clones and restriction maps on genomic DNA extracted from peripheral lymphocytes. Interestingly, only one example of restriction fragment length polymorphism was found during this analysis: a BglII polymorphism upstream from the HOX1H homeobox. We have now extended this analysis and isolated new cosmid clones (Fig. 1). Every cosmid clone overlaps for at least 7 kb adjacent clones. Homeobox sequences were identified using the oligonucleotide 5'TGGTTCCAGAACCGGCGGATGAA3' representing the most conserved portion of human homeoboxes. DNA fragments of interest were subcloned and nucleotide sequences were determined according to Sanger et al. (15).

Results

We have previously reported partial genomic maps of the HOX loci (16,17,9) and their chromosomal assignment (18). We have now extended the chromosome walking around the reported genomic regions by the screening of a genomic library constructed in the cosmid vector pcos2EMBL (10). Genomic maps obtained from isolated clones were confirmed by restriction analysis of human DNA extracted from peripheral lymphocytes. Homeoboxes were identified using a synthetic oligonucleotide representing the most conserved portion of the class I homeobox (see Materials and Methods), subcloned and sequenced.

Fig. 1. Genomic organization of mapped regions of the human HOX loci. Identified homeobox sequences are shown as filled boxes. Transcription is from left to right. Chromosomal localization is indicated in brackets after each HOX designation. Maps of various loci derive from the analysis of overlapping cosmid clones reported below them. E=EcoRI.

Fig. 1 shows such genomic maps of four regions of the HOX loci containing 30 homeoboxes. These regions span 90 Kb, 180 Kb, 160 Kb and 70 Kb of DNA of HOX1, 2, 3 and 4, respectively, including 8, 9, 7 and 6 homeoboxes. It is impossible to exclude for the moment that additional less-conserved homeobox sequences might be present interspersed among the conserved homeoboxes so far localized. Complete sequencing of the entire regions is required to settle this point.

The recommended nomenclature for human homeoboxes (9) is used throughout. Accordingly, HOX2A, 2B, 2C .... 2G, for example, are the human homologues of murine Hox-2.1, 2.2, 2.3 .... 2.7 and HOX4A is the homologue of Hox-4.1. Because all HOX homeoboxes so far analysed in mouse and man seem to belong to only four HOX loci, we attributed the murine Hox-5.1, 5.2 and 5.3 homeoboxes to the HOX4 locus and termed their human cognates HOX4B, 4C and 4D, respectively. The human Hox-5.4 homeobox recently reported (19) was termed HOX4E. The general scheme of this correspondence is shown in Fig. 4 below.

Fig. 2A shows the nucleotide sequence of 10 new homeoboxes and Fig. 2B the conceptual translation of all 30 human homeoboxes identified so far. The sequences of 20 of these homeoboxes were previously reported by us (9). Identification of HOX4C and 4E homeoboxes was also reported (19). Their published sequences are identical to those we report here.

It has been previously noticed that class I HDs may be grouped in sub-classes or groups on the basis of the primary peptide sequence (17,9,7). These groups may include human and mouse as well as frog, rat, zebrafish, salmon, chicken and

A

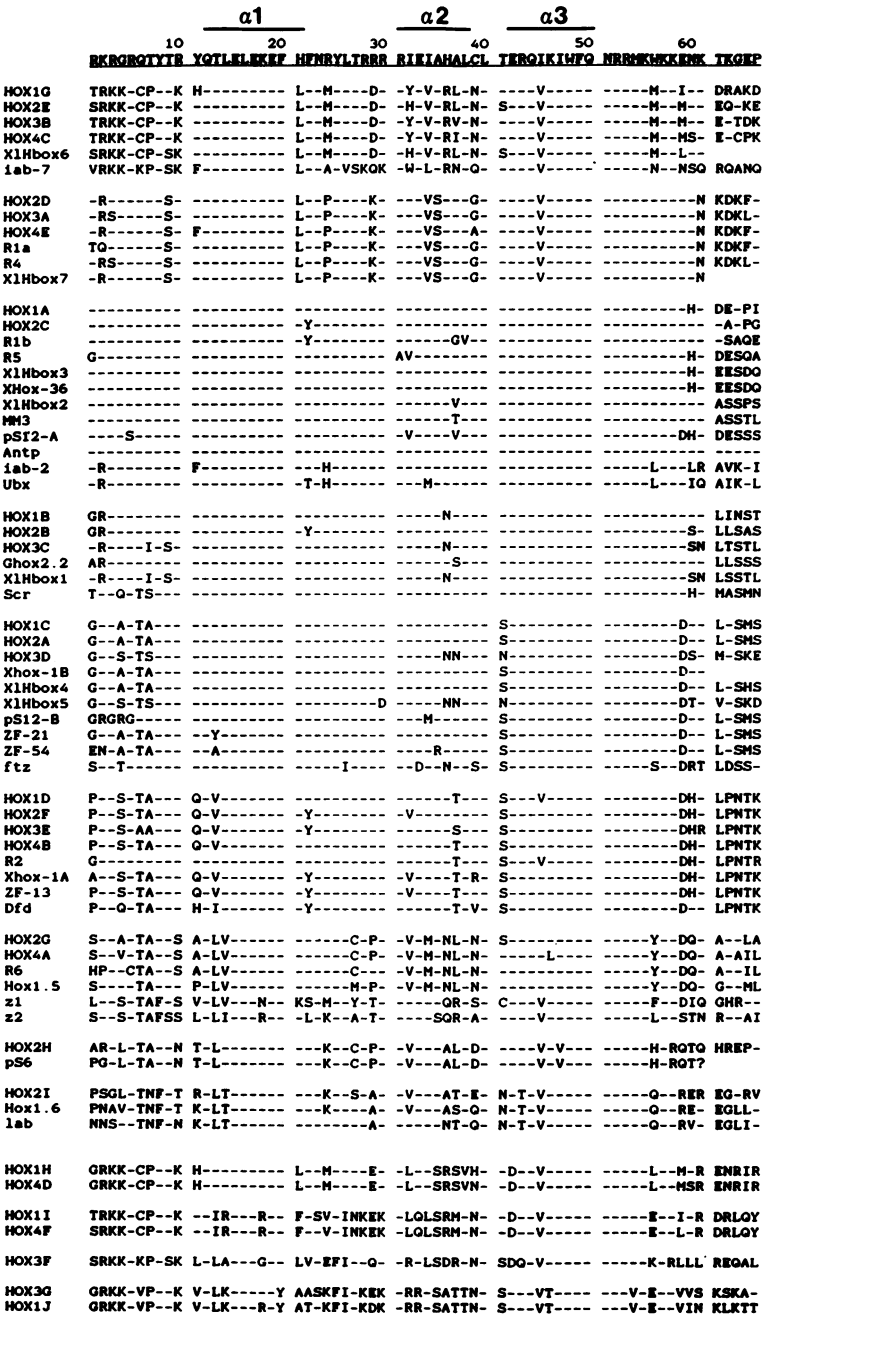
```
                                30                              60                              90
HOX1G   ACT CGG AAA AAG CGG TGC CCC TAT ACA AAA  CAC CAG ACC CTG GAA CTG GAG AAA GAG TTT  CTG TTC AAC ATG TAC CTC ACC AGG GAC CGC
HOX1H   GGT CGG AAG AAG CGC TGC CCC TAC ACG AAG  CAC CAG ACA CTG GAG CTG GAG AAG GAG TTT  CTG TTC AAT ATG TAC CTT ACT CGA GAG CGG
HOX1I   ACC CGC AAA AAG CGC TGC CCC TAT ACC AAG  TAC CAG ATC CGA GAG CTG GAA CGG GAG TTC  TTC TTC ACC GTC TAC ATT AAC AAA GAG AAG
HOX1J   GGG AGA AAG AAG CGC GTG CCT TAT ACC AAG  GTG CAA TTA AAA GAA CTT GAA CGG GAA TAC  GCC ACG AAT AAA TTC ATT ACT AAG GAC AAA
HOX2F   TCT CGG AAG AAG CGC AAG CCC TAT TCG AAG  TTG CAA CTG GCA GAG CTG GAG GCC GAG TTT  CTG GTC AAC GAG TTC ATC ACA CGC CAG CGC
HOX3G   GGG CGC AAG AAA CGC GTG CCC TAC ACT AAG  GTG CAG CTG AAG GAG CTA GAG AAG GAA TAC  GCG GCT AGC AAG TTC ATC ACC AAA GAG AAG
HOX4C   ACC CGC AAA AAG CGC TGT CCC TAC ACC AAA  TAC CAG ACC CTT GAG CTG GAG AAA GAA TTC  CTG TTC AAC ATG TAC CTC ACC CGG GAC CGC
HOX4D   GCC AGA AAG AAG AGG TGC CCT TAC ACT AAG  CAC CAA ACG CTG GAA TTA GAA AAA GAG TTC  TTG TTC AAT ATG TAC CTC ACC CGC GAG CGC
HOX4E   AGA CGC AGA CGA AGA CAA ACC TAC AGT CGC  TTC CAA ACT CTA GAG TTG GAA AAG GAA TTT  CTT TTT AAC CCC TAT CTG ACC AGG AAA AGA
HOX4F   TCC CGG AAA AAG CGC TGT CCC TAT ACC AAG  TAC CAG ATC CGC GAA CTG GAA CGC GAG TTT  TTC TTT AAC GTG TAC ATA AAC AAA GAG AAA

                               120                             150                             180
HOX1G   ACC TAC GAG GTG GCT CGA CTG CTC AAC CTC  ACC GAG AGG CAG GTC AAG ATC TGG TTC CAG  AAC CGC AGG ATG AAA ATG AAG AAA ATC AAC AAA
HOX1H   CGC CTA GAG ATT AGC CGC AGC GTC CAC CTC  ACG GAC AGA CAA GTG AAA ATC TGG TTT CAG  AAC CGC AGG ATG AAA CTG AAG AAA ATG AAT CGA
HOX1I   CGC CTG CAA CTG TCC CGC ATG CTC AAC CTC  ACT GAT CGT CAA GTC AAA ATC TGG TTT CAG  AAC AGG AGA ATG AAG GAA AAA AAA ATT AAC AGA
HOX1J   CGG AGG CGG ATA TCA GCC ACG ACG AAT CTC  TCT GAG CGG CAG GTC ACA ATC TGG TTC CAG  AAC AGG AGG GTT AAA GAG AAA AAA GTC ATC CAA
HOX2F   CGG AGG GAA CTC TCA GAC CGC TTG AAT CTT  AGT GAC CAG CAG GTC AAG ATC TGG TTT CAG  AAC CGG AGA ATG AAA AAG AAA AGA CTT CTG TTG
HOX3G   CGC CGG CGC ATC TCC GCC ACC ACG AAC CTC  TCT GAG CGC CAG GTA ACC ATC TGG TTC CAG  AAC CGG CGG GTC AAA GAG AAG AAG GTG GTC AGC
HOX4C   CGC TAC GAG GTG GCC AGG ATT CTC AAC CTA  ACA GAG AGA CAG GTG AAA ATC TGG TTT CAG  AAC CGT AGC ATG AAA ATG AAA AAG ATG AGC AAG
HOX4D   CGC CTA GAG ATC AGT AAG AGC GTT AAC CTC  ACC GAC AGG CAG GTC AAG ATT TGG TTT CAA  AAC CGC CGA ATG AAA CTC AAG AAG ATG AGC CGA
HOX4E   AGA ATC GAG GTT TCC CAC GCC CTA GCC CTC  ACC GAG AGA CAG GTA AAA ATC TGG TTC CAG  AAC AGG AGA ATG AAA TGG AAA AAG GAA AAC AAC
HOX4F   AGA CTT CAA CTC TCT CGG ATG CTC AAC CTC  ACT GAC CGG CAA GTC AAA ATC TGG TTC CAG  AAT CGC AGG ATG AAA GAA AAG AAA CTG AAC AGA
```

B

	10	20	30
HOX1A	Arg Lys Arg Gly Arg Gln Thr Tyr Thr Arg	Tyr Gln Thr Leu Glu Leu Glu Lys Glu Phe	His Phe Asn Arg Tyr Leu Thr Arg Arg Arg
HOX1B	Gly Arg Arg Gly Arg Gln Thr Tyr Thr Arg	Tyr Gln Thr Leu Glu Leu Glu Lys Glu Phe	His Phe Asn Arg Tyr Leu Thr Arg Arg Arg
HOX1C	Gly Lys Arg Ala Arg Thr Ala Tyr Thr Arg	Tyr Gln Thr Leu Glu Leu Glu Lys Glu Phe	His Phe Asn Arg Tyr Leu Thr Arg Arg Arg
HOX1D	Pro Lys Arg Ser Arg Thr Ala Tyr Thr Arg	Gln Gln Val Leu Glu Leu Glu Lys Glu Phe	His Phe Asn Arg Tyr Leu Thr Arg Arg Arg
HOX1G	Thr Arg Lys Lys Arg Cys Pro Tyr Thr Lys	His Gln Thr Leu Glu Leu Glu Lys Glu Phe	Leu Phe Asn Met Tyr Leu Thr Arg Asp Arg
HOX1H	Gly Arg Lys Lys Arg Cys Pro Tyr Thr Lys	His Gln Thr Leu Glu Leu Glu Lys Glu Phe	Leu Phe Asn Met Tyr Leu Thr Arg Glu Arg
HOX1I	Thr Arg Lys Lys Arg Cys Pro Tyr Thr Lys	Tyr Gln Ile Arg Glu Leu Glu Arg Glu Phe	Phe Phe Ser Val Tyr Ile Asn Lys Glu Lys
HOX1J	Gly Arg Lys Lys Arg Val Pro Tyr Thr Lys	Val Gln Leu Lys Glu Leu Glu Arg Glu Tyr	Ala Thr Asn Lys Phe Ile Thr Lys Asp Lys
HOX2A	Gly Lys Arg Ala Arg Thr Ala Tyr Thr Arg	Tyr Gln Thr Leu Glu Leu Glu Lys Glu Phe	His Phe Asn Arg Tyr Leu Thr Arg Arg Arg
HOX2B	Thr Ala Arg Gly Arg Gln Thr Tyr Thr Arg	Tyr Gln Thr Leu Glu Leu Glu Lys Glu Phe	His Tyr Asn Arg Tyr Leu Thr Arg Arg Arg
HOX2C	Arg Lys Arg Gly Arg Gln Thr Tyr Thr Arg	Tyr Gln Thr Leu Glu Leu Glu Lys Glu Phe	His Tyr Asn Arg Tyr Leu Thr Arg Arg Arg
HOX2D	Arg Arg Arg Gly Arg Gln Thr Tyr Ser Arg	Tyr Gln Thr Leu Glu Leu Glu Lys Glu Phe	Leu Phe Asn Pro Tyr Leu Thr Arg Lys Arg
HOX2E	Ser Arg Lys Lys Arg Cys Pro Tyr Thr Lys	Tyr Gln Thr Leu Glu Leu Glu Lys Glu Phe	Leu Phe Asn Met Tyr Leu Thr Arg Asp Arg
HOX2F	Pro Lys Arg Ser Arg Thr Ala Tyr Thr Arg	Gln Gln Val Leu Glu Leu Glu Lys Glu Phe	His Tyr Asn Arg Tyr Leu Thr Arg Arg Arg
HOX2G	Ser Lys Arg Ala Arg Thr Ala Tyr Thr Ser	Ala Gln Leu Val Glu Leu Glu Lys Glu Phe	His Phe Asn Arg Tyr Leu Cys Arg Pro Arg
HOX2H	Ala Arg Arg Leu Arg Thr Ala Tyr Thr Asn	Thr Gln Leu Leu Glu Leu Glu Lys Glu Phe	His Phe Asn Lys Tyr Leu Cys Arg Pro Arg
HOX2I	Pro Ser Gly Leu Arg Thr Asn Phe Thr Thr	Arg Gln Leu Thr Glu Leu Glu Lys Glu Phe	His Phe Asn Lys Tyr Leu Ser Arg Ala Arg
HOX3A	Arg Arg Ser Gly Arg Gln Thr Tyr Ser Arg	Tyr Gln Thr Leu Glu Leu Glu Lys Glu Phe	Leu Phe Asn Pro Tyr Leu Thr Arg Lys Arg
HOX3B	Thr Arg Lys Lys Arg Cys Pro Tyr Thr Lys	Tyr Gln Thr Leu Glu Leu Glu Lys Glu Phe	Leu Phe Asn Met Tyr Leu Thr Arg Asp Arg
HOX3C	Arg Arg Arg Gly Arg Gln Ile Tyr Ser Arg	Tyr Gln Thr Leu Glu Leu Glu Lys Glu Phe	His Phe Asn Arg Tyr Leu Thr Arg Arg Arg
HOX3D	Gly Lys Arg Ser Arg Thr Ser Tyr Thr Arg	Tyr Gln Thr Leu Glu Leu Glu Lys Glu Phe	His Phe Asn Arg Tyr Leu Thr Arg Arg Arg
HOX3E	Pro Lys Arg Ser Arg Ala Ala Tyr Thr Arg	Gln Gln Val Leu Glu Leu Glu Lys Glu Phe	His Tyr Asn Arg Tyr Leu Thr Arg Arg Arg
HOX3F	Ser Arg Lys Lys Arg Lys Pro Tyr Ser Lys	Leu Gln Leu Ala Glu Leu Glu Gly Glu Phe	Leu Val Asn Glu Phe Ile Thr Arg Gln Arg
HOX3G	Gly Arg Lys Lys Arg Val Pro Tyr Thr Lys	Val Gln Leu Lys Glu Leu Glu Lys Glu Tyr	Ala Ala Ser Lys Phe Ile Thr Lys Glu Lys
HOX4A	Ser Lys Arg Val Arg Thr Thr Tyr Thr Ser	Ala Gln Leu Val Glu Leu Glu Lys Glu Phe	His Phe Asn Arg Tyr Leu Cys Arg Pro Arg
HOX4B	Pro Lys Arg Ser Arg Thr Ala Tyr Thr Arg	Gln Gln Val Leu Glu Leu Glu Lys Glu Phe	His Phe Asn Arg Tyr Leu Thr Arg Arg Arg
HOX4C	Thr Arg Lys Lys Arg Cys Pro Tyr Thr Lys	Tyr Gln Thr Leu Glu Leu Glu Lys Glu Phe	Leu Phe Asn Met Tyr Leu Thr Arg Asp Arg
HOX4D	Gly Arg Lys Lys Arg Cys Pro Tyr Thr Lys	His Gln Thr Leu Glu Leu Glu Lys Glu Phe	Leu Phe Asn Met Tyr Leu Thr Arg Glu Arg
HOX4E	Arg Arg Arg Gly Arg Gln Thr Tyr Ser Arg	Phe Gln Thr Leu Glu Leu Glu Lys Glu Phe	Leu Phe Asn Pro Tyr Leu Thr Arg Lys Arg
HOX4F	Ser Arg Lys Lys Arg Cys Pro Tyr Thr Lys	Tyr Gln Ile Arg Glu Leu Glu Arg Glu Phe	Phe Phe Asn Val Tyr Ile Asn Lys Glu Lys

	40	50	60
HOX1A	Arg Ile Glu Ile Ala His Ala Leu Cys Leu	Thr Glu Arg Gln Ile Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Trp Lys Lys Glu His Lys
HOX1B	Arg Ile Glu Ile Ala Asn Ala Leu Cys Leu	Thr Glu Arg Gln Ile Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Trp Lys Lys Glu Asn Lys
HOX1C	Arg Ile Glu Ile Ala His Ala Leu Cys Leu	Ser Glu Arg Gln Ile Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Trp Lys Lys Asp Asn Lys
HOX1D	Arg Ile Glu Ile Ala His Thr Leu Cys Leu	Ser Glu Arg Gln Val Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Trp Lys Lys Asp His Lys
HOX1G	Arg Tyr Glu Val Ala Arg Leu Leu Asn Leu	Thr Glu Arg Gln Val Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Met Lys Lys Ile Asn Lys
HOX1H	Arg Leu Glu Ile Ser Arg Ser Val His Leu	Thr Asp Arg Gln Val Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Leu Lys Lys Met Asn Arg
HOX1I	Arg Leu Gln Leu Ser Arg Met Leu Asn Leu	Thr Asp Arg Gln Val Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Glu Lys Lys Ile Asn Arg
HOX1J	Arg Arg Arg Ile Ser Ala Thr Thr Asn Leu	Ser Glu Arg Gln Val Thr Ile Trp Phe Gln	Asn Arg Arg Val Lys Glu Lys Lys Val Ile Asn
HOX2A	Arg Ile Glu Ile Ala His Ala Leu Cys Leu	Ser Glu Arg Gln Ile Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Trp Lys Lys Asp Asn Lys
HOX2B	Arg Ile Glu Ile Ala His Ala Leu Cys Leu	Thr Glu Arg Gln Ile Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Trp Lys Lys Glu Ser Lys
HOX2C	Arg Ile Glu Ile Ala His Ala Leu Cys Leu	Thr Glu Arg Gln Ile Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Trp Lys Lys Glu Asn Lys
HOX2D	Arg Ile Glu Val Ser His Ala Leu Gly Leu	Thr Glu Arg Gln Val Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Trp Lys Lys Glu Asn Asn
HOX2E	Arg His Glu Val Ala Arg Leu Leu Asn Leu	Ser Glu Arg Gln Val Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Met Lys Lys Met Asn Lys
HOX2F	Arg Val Glu Ile Ala His Ala Leu Cys Leu	Ser Glu Arg Gln Ile Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Trp Lys Lys Asp His Lys
HOX2G	Arg Val Glu Met Ala Asn Leu Leu Asn Leu	Ser Glu Arg Gln Ile Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Tyr Lys Lys Asp Gln Lys
HOX2H	Arg Val Glu Ile Ala Ala Leu Leu Asp Leu	Thr Glu Arg Gln Val Lys Val Trp Phe Gln	Asn Arg Arg Met Lys His Lys Arg Gln Thr Gln
HOX2I	Arg Val Glu Ile Ala Ala Thr Leu Glu Leu	Asn Glu Thr Gln Val Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Gln Lys Lys Arg Glu Arg
HOX3A	Arg Ile Glu Val Ser His Ala Leu Gly Leu	Thr Glu Arg Gln Val Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Trp Lys Lys Glu Asn Asn
HOX3B	Arg Tyr Glu Val Ala Arg Val Leu Asn Leu	Thr Glu Arg Gln Val Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Met Lys Lys Met Asn Lys
HOX3C	Arg Ile Glu Ile Ala Asn Ala Leu Cys Leu	Thr Glu Arg Gln Ile Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Trp Lys Lys Glu Ser Asn
HOX3D	Arg Ile Glu Ile Ala Asn Asn Leu Cys Leu	Asn Glu Arg Gln Ile Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Trp Lys Lys Asp Ser Lys
HOX3E	Arg Ile Glu Ile Ala His Ser Leu Cys Leu	Ser Glu Arg Gln Ile Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Trp Lys Lys Asp His Arg
HOX3F	Arg Arg Glu Leu Ser Asp Arg Leu Asn Leu	Ser Asp Gln Gln Val Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Lys Lys Arg Leu Leu Leu
HOX3G	Arg Arg Arg Ile Ser Ala Thr Thr Asn Leu	Ser Glu Arg Gln Val Thr Ile Trp Phe Gln	Asn Arg Arg Val Lys Glu Lys Lys Val Val Ser
HOX4A	Arg Val Glu Met Ala Asn Leu Leu Asn Leu	Thr Glu Arg Gln Ile Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Tyr Lys Lys Asp Gln Lys
HOX4B	Arg Ile Glu Ile Ala His Thr Leu Cys Leu	Ser Glu Arg Gln Ile Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Trp Lys Lys Asp His Lys
HOX4C	Arg Tyr Glu Val Ala Arg Ile Leu Asn Leu	Thr Glu Arg Gln Val Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Met Lys Lys Met Ser Lys
HOX4D	Arg Leu Glu Ile Ser Lys Ser Val Asn Leu	Thr Asp Arg Gln Val Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Leu Lys Lys Met Ser Arg
HOX4E	Arg Ile Glu Val Ser His Ala Leu Ala Leu	Thr Glu Arg Gln Val Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Trp Lys Lys Glu Asn Asn
HOX4F	Arg Leu Gln Leu Ser Arg Met Leu Asn Leu	Thr Asp Arg Gln Val Lys Ile Trp Phe Gln	Asn Arg Arg Met Lys Glu Lys Lys Leu Asn Arg

Fig. 2. Identified human HOX homeoboxes.
A) Nucleotide sequences of 10 new homeoboxes.
B) Alignment of homeodomains obtained from the conceptual translation of homeobox sequences shown in 2A) and 20 previously reported (9) homeodomains.


α1
α2
α3
10
20
30
40
50
60
RKRGRQTYTR YQTLELEKEF HFNRYLTRRR RIEIAHALCL TERQIKIWFQ NRRMKWKKEN TKGKP
HOX1G TRKK-CP--K H--------- L--M----D- -Y-V-RL-N- ----V----- -----M--I-- DRAKD
HOX2E SRKK-CP--K ---------- L--M----D- -H-V-RL-N- S---V----- -----M--M-- EQ-KE
HOX3B TRKK-CP--K ---------- L--M----D- -Y-V-RV-N- ----V----- -----M--M-- E-TDK
HOX4C TRKK-CP--K ---------- L--M----D- -Y-V-RI-N- ----V----- -----M--MS- E-CPK
Xlhbox6 SRKK-CP-SK ---------- L--M----D- -H-V-RL-N- S---V----- -----M--L--
iab-7 VRKK-KP-SK F--------- L--A-VSKQK -W-L-RN-Q- ----V----- -----N--NSQ RQANQ
HOX2D -R------S- ---------- L--P----K- ---VS---G- ----V----- ----------N KDKF-
HOX3A -RS-----S- ---------- L--P----K- ---VS---G- ----V----- ----------N KDKL-
HOX4E -R------S- F--------- L--P----K- ---VS---A- ----V----- ----------N KDKF-
R1a TQ------S- ---------- L--P----K- ---VS---G- ----V----- ----------N KDKF-
R4 -RS-----S- ---------- L--P----K- ---VS---G- ----V----- ----------N KDKL-
Xlhbox7 -R------S- ---------- L--P----K- ---VS---G- ----V----- ----------N
HOX1A ---------- ---------- ---------- ---------- ---------- ---------H- DE-PI
HOX2C ---------- ---------- -Y-------- ---------- ---------- ---------- -A-PG
R1b ---------- ---------- -Y-------- ------GV-- ---------- ---------- -SAQE
R5 G--------- ---------- ---------- AV-------- ---------- ---------H- DESQA
Xlhbox3 ---------- ---------- ---------- ---------- ---------- ---------H- EESDQ
XHox-36 ---------- ---------- ---------- ---------- ---------- ---------H- EESDQ
Xlhbox2 ---------- ---------- ---------- ------V--- ---------- ---------- ASSPS
MM3 ---------- ---------- ---------- ------T--- ---------- ---------- ASSTL
pSI2-A ----S----- ---------- ---------- -V----V--- ---------- --------DH- DESSS
Antp ---------- ---------- ---------- ---------- ---------- ---------- -----
iab-2 -R-------- F--------- ---H------ ---------- ---------- -----L---LR AVK-I
Ubx -R-------- ---------- -T-H------ ---M------ ---------- -----L---IQ AIK-L
HOX1B GR-------- ---------- ---------- -----N---- ---------- ---------- LINST
HOX2B GR-------- ---------- -Y-------- ---------- ---------- ---------S- LLSAS
HOX3C -R----I-S- ---------- ---------- -----N---- ---------- ---------SN LTSTL
Ghox2.2 AR-------- ---------- ---------- ------S--- ---------- ---------- LLSSS
Xlhbox1 -R----I-S- ---------- ---------- -----N---- ---------- ---------SN LSSTL
Scr T--Q-TS--- ---------- ---------- ---------- ---------- ---------H- MASMN
HOX1C G--A-TA--- ---------- ---------- ---------- S--------- --------D-- L-SMS
HOX2A G--A-TA--- ---------- ---------- ---------- S--------- --------D-- L-SMS
HOX3D G--S-TS--- ---------- ---------- -----NN--- N--------- --------DS- M-SKE
Xhox-1B G--A-TA--- ---------- ---------- ---------- S--------- --------D--
Xlhbox4 G--A-TA--- ---------- ---------- ---------- S--------- --------D-- L-SMS
Xlhbox5 G--S-TS--- ---------- ---------D ------NN--- N--------- --------DT- V-SKD
pSI2-B GRGRG----- ---------- ---------- ---M------ S--------- --------D-- L-SMS
ZF-21 G--A-TA--- --Y------- ---------- ---------- S--------- --------D-- L-SMS
ZF-54 EN-A-TA--- --A------- ---------- ----R----- S--------- --------D-- L-SMS
ftz S--T------ ---------- -----I----- --D--N--S- S--------- -----S--DRT LDSS-
HOX1D P--S-TA--- Q-V------- ---------- ------T--- S---V----- --------DH- LPNTK
HOX2F P--S-TA--- Q-V------- -Y-------- -V-------- S--------- --------DH- LPNTK
HOX3E P--S-AA--- Q-V------- -Y-------- ------S--- S--------- --------DHR LPNTK
HOX4B P--S-TA--- Q-V------- ---------- ------T--- S--------- --------DH- LPNTK
R2 G--------- ---------- ---------- ------T--- S---V----- --------DH- LPNTR
Xhox-1A A--S-TA--- Q-V------- -Y-------- -V----T-R- S--------- --------DH- LPNTK
ZF-13 P--S-TA--- Q-V------- -Y-------- -V----T--- S--------- --------DH- LPNTK
Dfd P--Q-TA--- H-I------- -Y-------- ------T-V- S--------- --------D-- LPNTK
HOX2G S--A-TA--S A-LV------ ------C-P- -V-M-NL-N- S--------- -----Y--DQ- A--LA
HOX4A S--V-TA--S A-LV------ ------C-P- -V-M-NL-N- -----L---- -----Y--DQ- A-AIL
R6 HP--CTA--S A-LV------ ------C--- -V-M-NL-N- ---------- -----Y--DQ- A--IL
Hox1.5 S----TA--- P-LV------ ------M-P- -V-M-NL-N- ---------- -----Y--DQ- G--ML
z1 L--S-TAF-S V-LV---N-- KS-M--Y-T- -----QR-S- C---V----- -----F--DIQ GHR--
z2 S--S-TAFSS L-LI---R-- -L-K--A-T- ----SQR-A- ----V----- -----L--STN R--AI
HOX2H AR-L-TA--N T-L------- ---K--C-P- -V---AL-D- ----V-V--- -----H-RQTQ HREP-
pS6 PG-L-TA--N T-L------- ---K--C-P- -V---AL-D- ----V-V--- -----H-RQT?
HOX2I PSGL-TNF-T R-LT------ ---K--S-A- -V---AT-E- N-T-V----- -----Q--RER EQ-RV
Hox1.6 PNAV-TNF-T K-LT------ ---K----A- -V---AS-Q- N-T-V----- -----Q--RE- EGLL-
lab NNS--TNF-N K-LT------ --------A- -----NT-Q- N-T-V----- -----Q--RV- EGLI-
HOX1H GRKK-CP--K H--------- L--M----E- -L--SRSVH- -D--V----- -----L--M-R ENRIR
HOX4D GRKK-CP--K H--------- L--M----E- -L--SRSVN- -D--V----- -----L--MSR ENRIR
HOX1I TRKK-CP--K --IR---R-- F-SV-INKEK -LQLSRM-N- -D--V----- -----E--I-R DRLQY
HOX4F SRKK-CP--K --IR---R-- F--V-INKEK -LQLSRM-N- -D--V----- -----E--L-R DRLQY
HOX3F SRKK-KP-SK L-LA---G-- LV-EFI--Q- -R-LSDR-N- SDQ-V----- -----K-RLLL REQAL
HOX3G GRKK-VP--K V-LK-----Y AASKFI-KEK -RR-SATTN- S---VT---- ---V-E--VVS KSKA-
HOX1J GRKK-VP--K V-LK---R-Y AT-KFI-KDK -RR-SATTN- S---VT---- ---V-E--VIN KLKTT

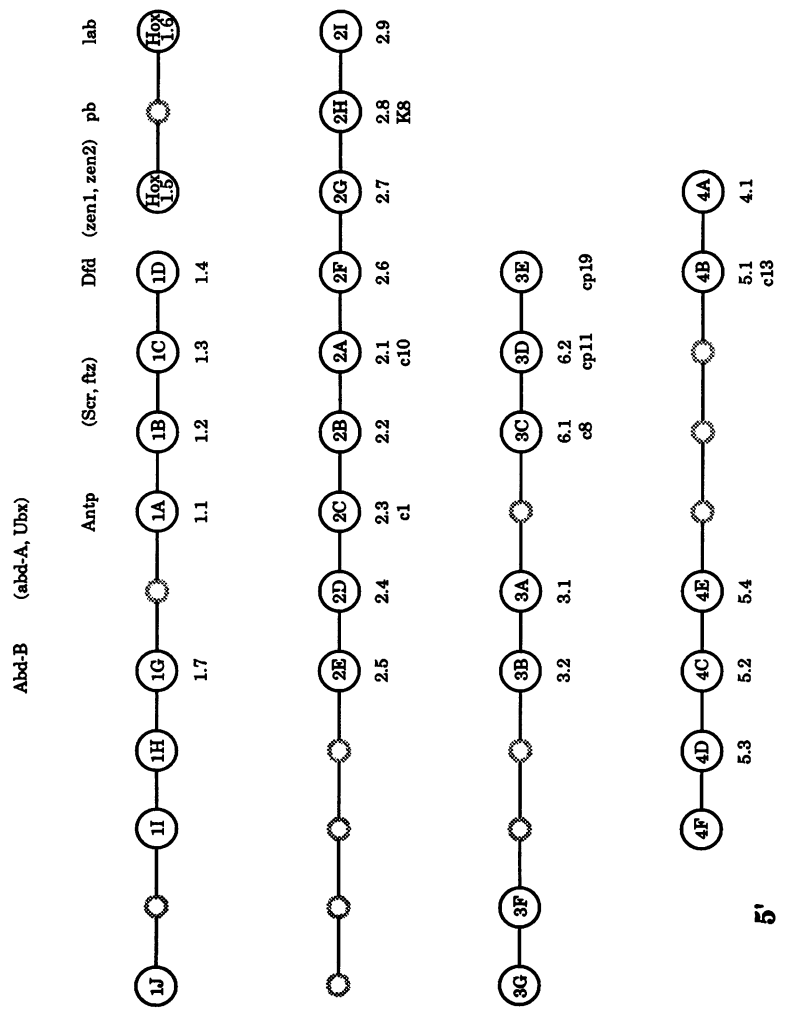
Drosophila HDs. For example, HOX1A, 2C and frog MM3 HDs differ by a single amino acid residue from the Drosophila Antp HD. These four HDs identify a sub-class or HD group. HD groups thus obtained are shown in Fig. 3.

Nine groups were previously reported (9,20). Seven new human homeoboxes have now been identified upstream from the HD group containing HOX1G, 2E, 3B and 4C. On the basis of their primary sequence the encoded HDs appear to belong to at least four additional groups (Fig. 3), the first comprising HOX1H and 4D (homologous to the murine Hox-5.3 (21)), the second HOX1I and 4F, the third HOX3F and the fourth HOX3G and possibly HOX1J. It is noteworthy that HOX1H and 4D share also the pentapeptide Glu-Asn-Arg-Ile-Arg immediately downstream from the HD (Fig. 3). Similarly, HOX1I and 4F share the downstream pentapeptide Asp-Arg-Leu-Gln-Tyr. It has already been noticed that often the HDs of the same group share homologous downstream pentapeptides, most notably in the group including HOX1D, 2F, 3E and 4B and the Deformed (Dfd) HDs (17,9).

Therefore, at least 13 human HD groups can now be identified. Human HDs belonging to the same group occupy homologous positions in their respective HOX loci and it is possible to align the four HOX loci of Fig. 1. so that corresponding HDs in all loci share the maximal peptide sequence identity (17,9). For example, HDs 2A and 1C are identical, HDs 2D and 3A differ by a single amino acid residue, HD 3D differs for 6 residues from HD1C but for 9 residues from both HDs 1B and 1D and so on (Figs. 2B and 3). The alignment of human loci thus obtained is shown in a schematic form in Fig. 4, where alternative designations of reported HDs have been indicated. Murine Hox-1.5 and 1.6 HDs have also been included to complete the picture as their human homologues have not yet been isolated.

The correspondence of individual HDs in the four HOX loci suggests the hypothesis of large-scale duplications of a single homeobox gene complex with a subsequent dispersion in different chromosomes (22,17,9). Obviously it is not necessary for every HOX locus to contain the same number of homeoboxes. In fact, no murine homeobox has been identified between Hox-1.5 and Hox-1.6 and for the time being we have failed to detect predicted homeoboxes between HOX1G and 1A, between HOX3A and 3C and HOX4E

Fig. 3. List of homeodomains.
The one-letter amino acid code is used. All sequences are compared to the Antp HD. Dashes indicate amino acid identity. Human HOX and 2 murine HDs (Hox-1.5 and Hox-1.6) are named according to accepted nomenclature. Other HDs retain their original identification. Grouped sequences represent closely related HDs. The first 9 groupings correspond to groups I-IX (9) of Fig. 4 below. Other 4 tentative human groupings follow. 5 amino acid residues following the HD are also shown. XlHbox 1 to 7, MM3, Xhox-36, Xhox-1A and -1B are Xenopus HDs; R designates rat HDs; pS are salmon HDs (34); ZF HDs are from zebrafish and Ghox2.2 is from chicken (35). All remaining HDs are from Drosophila genes. Sequences are from Ref.7 when not otherwise indicated. The three α-helix domains are indicated.


Abd-B (abd-A, Ubx)
Antp (Scr, ftz) Dfd (zen1, zen2) pb lab
chr 7
chr 17
chr 12
chr 2
5'
3'
Group I II III IV V VI VII VIII IX

and 4B. In the far upstream regions of the four HOX loci we failed to detect any HOX1 homeobox between HOX1J and 1I, any HOX2 homeobox upstream from 2E, and any HOX3 homeobox between HOX3F and 3B (Fig. 4). After the separation of the four loci some homeoboxes may have been lost because of accumulated mutations or by intrachromosomal deletions. Conversely, some homeoboxes might have been gained by gene duplication (17).

It has been shown that some mammalian HD groups can be put in a one-to-one correspondence with the HDs contained in Drosophila homeotic genes present in ANT-C and BX-C (17,20,21, 9). As indicated in Fig. 4, there is a particularly significant correspondence between mammalian groups and fly homeotic genes Abdominal-B (Abd-B), Antp, Dfd, proboscipedia (pb) and labial (lab), in this order, 5' to 3' with respect to the transcriptional orientation of the mammalian homeoboxes. These observations suggest that the mammalian HOX loci are true homologues of the insect homeotic gene complexes (reviewed in Ref. 2). Comparative analysis of vertebrate and fly HDs shows that amino acid changes within the α-helix subdomains are remarkably constant within individual groups and show an ordered variation between different groups (Fig. 3), suggesting that the selective pressure on the peptide sequences has allowed a moderate diversification of the various HD groups in specific structural subdomains (9). This ordered diversification must play a fundamental role in the regulatory network involving HOX and homeotic gene products.

On the other hand, the HD represents only a domain of these gene products and a comparative analysis of vertebrate and fly homeoproteins is necessary. Some predicted homeoproteins encoded by vertebrate and fly homeobox genes have been reported. We analysed the predicted gene products corresponding to human HOX2G, 2H and 2I (Figs. 5-8).

Using HOX2G, 2H and 2I homeoboxes as probes we screened a cDNA library prepared from poly(A)+ RNA of human teratocarcinoma N-TERA2 cells cultured for 14 days in 10 μM retinoic acid (12, 13). Three cDNA clones were isolated containing extended open reading frames including the expected HDs of HOX2G, 2H and 2I, respectively. Fig. 5 shows the intron-exon organization of the

Fig. 4. Schematic representation of 30 human HOX HDs (circles) in the 4 chromosomal loci. Below the circles are shown the names of known murine Hox homologues along with laboratory designations of 6 HOX HDs (c1, c8, c10, c13, cp11 and cp19). K8 designation is from Ref. 30. The four loci have been aligned in such a way as to minimize amino acid changes of HDs placed in the same column. Murine Hox-1.5 (22) and Hox-1.6 (25,28) HDs have also been included. Stippled small circles indicate HDs predicted in the scheme but not yet identified. Groups I-IX of Ref. 9 are shown below the scheme. HDs from Drosophila BX-C and ANT-C genes are indicated above the scheme. Each fly HD has been placed on top of the group of human HDs to which it is most closely related in sequence. Correspondence of HDs in brackets is unclear. The sequence of the pb HD has not been published, but it appears to differ from the HOX2H HD for only 4 amino acid changes (David, L. Cribbs and Thomas C. Kaufman, personal communication).

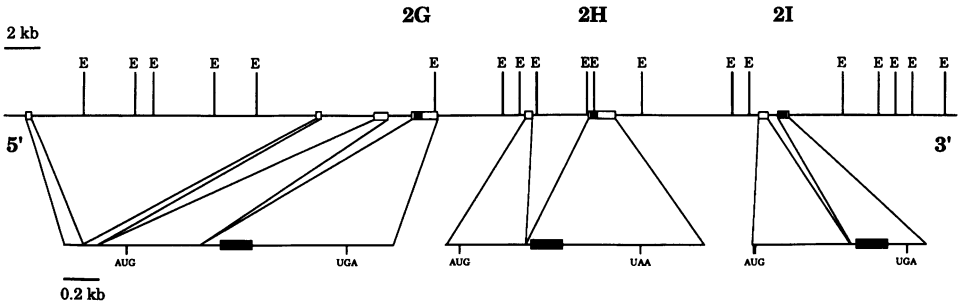


Fig. 5. Intron-exon organization of three cDNA clones containing HOX2G, 2H and 2I homeoboxes. The EcoRI (E) map is derived from clones CD2B and BC206 shown in Fig. 1. Boxes indicate exons and filled boxes represent homeoboxes. Initiation and stop codons are indicated in the detailed map of cDNA clones.

three cDNAs we have determined on overlapping cosmid clones CD2B and BC206 (Fig. 1). In particular, the HOX2G clone we isolated appears to be comprised of 4 exons spanning 25 kb of HOX2 DNA (Figs. 5 and 6), whereas the other two cDNAs contain 2 exons separated by a short intron (Figs. 5,7,8). In every case the homeobox is present in the 3' exon, as commonly reported so far.

Fig. 6 shows the sequence of the 1.9 Kb cDNA clone encoding the HOX2G homeoprotein of 431 amino acid residues. It contains a relatively long C-terminal region (183 residues) downstream from the HD and an unusually high number of amino acid residues separating the HD from the conserved pentapeptide upstream from it, variously called YPWM homeopeptide (24,33) or pre-box (23). This domain appears to be extended by the presence of a stretch of 23 Gly residues with two interspersed Ser residues. This unique peptide sequence is present in the HD-containing exon just downstream from the splice site. A comparison with the sequence of the related HOX4A homeoprotein reveals that this 25 residue motif is inserted in a precise site of an otherwise highly conserved peptide region spanning the splice site (Fig. 6).

Fig. 7 shows the sequence of a 1.5 Kb HOX2H cDNA clone encoding a 356 residue protein. Also this homeoprotein is characterized by a long 153-residue C-terminus and by a 44-residue domain lying between the HD and the YPWM homeopeptide. The peptide domain preceding the conserved pentapeptide is particularly Pro-rich with sequences of 5 and 7 proline residues in a row.

Finally, Fig. 8A shows a HOX2I coding region obtained by sequencing a short cDNA clone. Comparison of the predicted N-terminus with that of the murine Hox-1.6 homeoprotein (25) reveals that 5 amino acid residues upstream from the reported Hox-1.6 ATG are conserved between the two sequences. This homology suggests that 5 additional residues might belong to the coding region of the two homeoproteins. Alternatively, the

sequence homology may play a role in translational control. Also in this case the peptide region between the HD and the YPWM homeopeptide appears to be of interest (Fig. 8B). In fact, there is a remarkable degree of homology in this domain between the lab and HOX2I homeoproteins as well as between the lab and Hox-1.6 homeoproteins, as previously noticed (26).

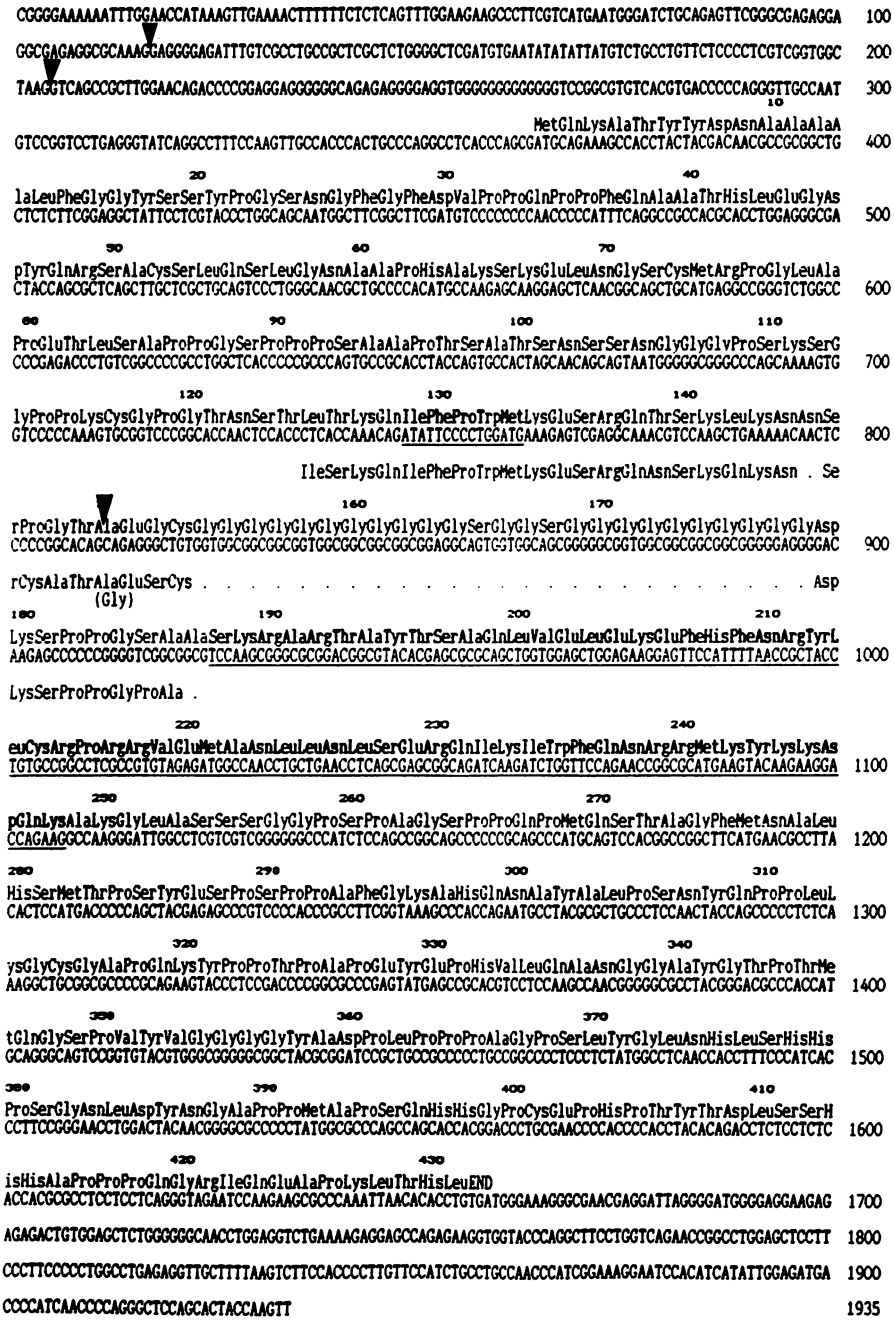
Discussion

Evidence is accumulating that the HOX loci in mouse (22) man (9) and probably Xenopus (27) arose as duplications of an ancestral complex locus containing several homeoboxes. Though the number of homeoboxes in the four loci is not strictly conserved their relative order is. The encoded HDs share a high degree of homology to each other but it is possible to assign every HD to a given sub-class or group by inspection of its primary sequence. HDs of the same sequence group occupy corresponding positions in the various HOX loci. This is in part due to the common origin of the four loci as duplications of an ancestral complex locus but we have shown (9) that a selective pressure on the peptide sequence has been necessary to maintain the ordered diversity of the various HDs within every locus. In fact, if one takes into consideration only base substitutions at the third position of the 61 codons it appears that all human homeoboxes are uniformly divergent from each other and from Drosophila homeoboxes (around 50%) (9).

It is by no means clear why several slightly different HDs have to be maintained in an ordered sequence within every locus. One can speculate that different HDs bind to different DNA targets. Alternatively, they may bind to identical or similar targets with different affinity. In fact, amino acid changes between different HD groups are not randomly distributed within the entire domain but are clustered according to what are believed to be functional sub-domains, like the 3 α-helixes and the ß-turn (9).

Mammalian HOX loci have been shown to be true homologues of the Drosophila homeotic gene complexes (reviewed in Ref.2). Some of the various HD groups were already distinct when lineages leading to insects and vertebrates diverged. This correspondence is particularly clear for the Drosophila HDs of Abd-B, Dfd, pb and lab and their vertebrate counterparts (Figs. 3 and 4). Identification of other one-to-one correspondences is a little more controversial (31,2,9, 20,21) but the overall picture is clear. Moreover, corresponding genes of vertebrate and Drosophila loci show the same relative boundaries of expression domains along the antero-posterior axis of the developing embryo. Mammalian and fly genes on the right hand site of Fig. 4 are expressed in anterior regions whereas genes on the left hand site are expressed in more and more posterior regions (2).

Present data confirm and extend this evolutionary scheme. We report the existence of 7 human homeoboxes upstream from the Abd-B-like group which we had previously called group I (17,9). These new homeoboxes are present in HOX1, 3 and 4 but not in HOX2. The encoded HDs are clearly related to the Abd-B HD,



especially in the first 10 amino acid residues and can be divided in 4 (or 5) HD groups (Fig. 3) bringing the current total number of groups to 13 (or 14). There is some uncertainty as to whether HOX1J and 3G belong to the same group. HOX1H and 4D HDs differ for one amino acid change from each other and for 21 and 20 changes, respectively, from Abd-B; HOX1I and 4F HDs differ for 3 changes from each other and for 23 and 22 changes from Abd-B; the HOX3FHD differs for 25 changes from Abd-B. On the other hand, HOX3G and 1J HDs differ for 6 changes from each other and both differ for 30 changes from Abd-B. In addition, whereas HOX1H and 4D share an identical pentapeptide downstream from their HD as is the case for HOX1I and 4F, HOX1J and 3G share only a limited homology in this domain (Fig. 3).

What is the evolutionary position of the new homeoboxes with regard to the Drosophila gene complexes? They were either present in the ancestral locus predating the divergence between insects and vertebrates or, alternatively, they have arisen specifically in the lineage leading to vertebrates. If one accepts the first hypothesis it is necessary to explain why they seem to be absent in the Drosophila complexes, in particular in BX-C. One possible explanation is that they are present in the Drosophila genome, clustered in a different gene complex or scattered. Another possibility is that they were lost in the evolutionary lineage leading to flies, due to some peculiarity in their development. As a matter of fact, terminal abdominal segments appear to be missing or fused in Drosophila as compared to the situation of other insects (31,32). A third possibility is more intriguing. The 5 mammalian upstream HD groups may all correspond functionally to Abd-B. In Drosophila a huge genomic region containing several elements (iab-3 to iab-7) controlling the expression of the Abd-B homeoprotein might play the role of 5 mammalian homeoproteins.

On the other hand, these new HD groups may have arisen specifically in the evolutionary lineage leading to vertebrates in view of the development of specific body structures or in the frame of a general increase of complexity. Specific body structures that could be considered are, for example, the hindlimbs and the pelvic girdle or genitalia. Drosophila legs are thoracic whereas hindlimbs of tetrapods imply very specialized anatomical structures localized posteriorly. In situ hybridization experiments are required to assess these hypotheses along with a genetical analysis possibly through reverse genetics and analysis of transgenics.

The HD is only a portion of the homeoproteins. Henceforth, comparison of mammalian and fly homeoproteins might cast

Fig. 6. Nucleotide sequence of a HOX2G cDNA clone and its conceptual translation. The conserved pentapeptide and the HD are underlined. Arrowheads point to splice sites. The peptide sequence of a domain of HOX4A has been aligned to the corresponding HOX2G domain. Dots indicate a deletion. A glycine residue present in HOX4A, but not in HOX2G, is shown in brackets.

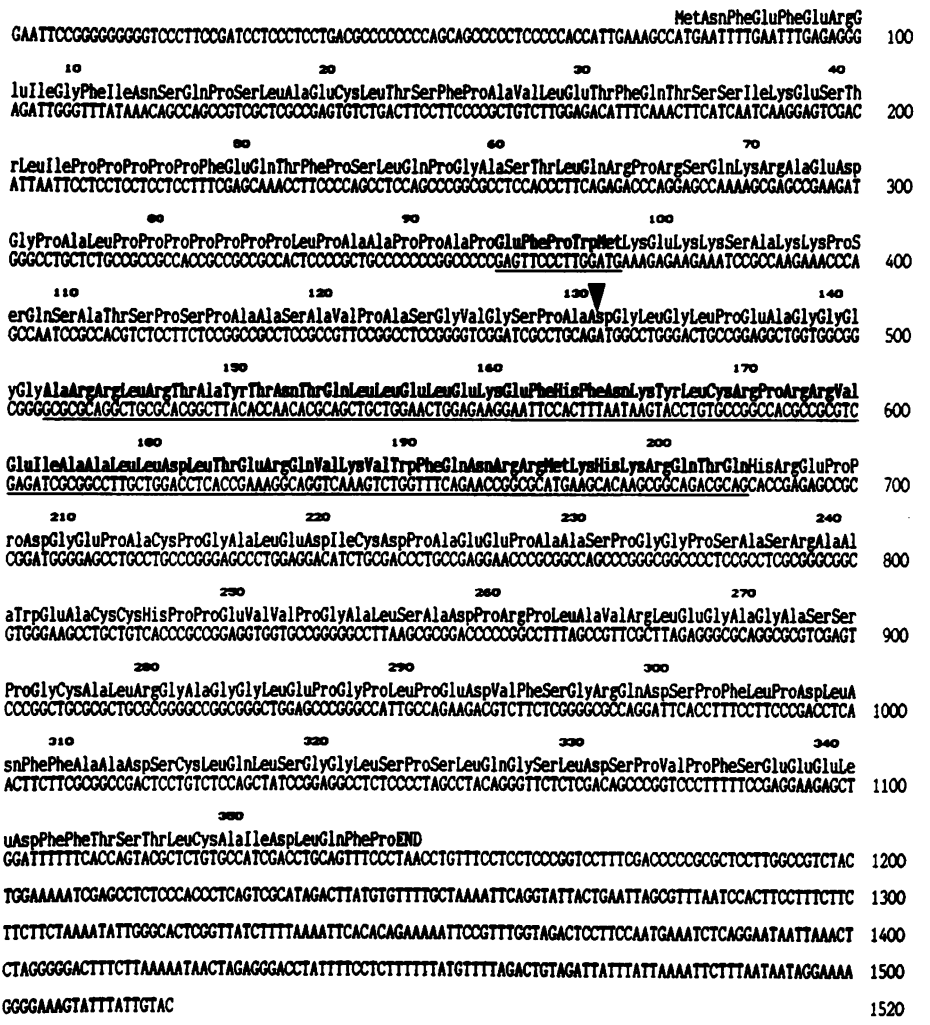


Fig. 7. A HOX2H cDNA clone. Symbols as in Fig. 6.

additional light on the evolution of these functional structures. Often the pentapeptide immediately downstream from the HD is conserved in human homeoproteins of the same group (Fig. 3). This is particularly evident for the Dfd-like group but also for groups lying both to the far right and far left of Fig. 4.

A second conserved motif in mammalian and fly homeoproteins is the pentapeptide centered around the YPWM core. We found it in every human HOX homeoprotein we looked at. They are 15 so far, namely HOX1B, 1D, 2A, 2C, 2D, 2F, 2G, 2H, 2I, 3A, 3C, 3D,

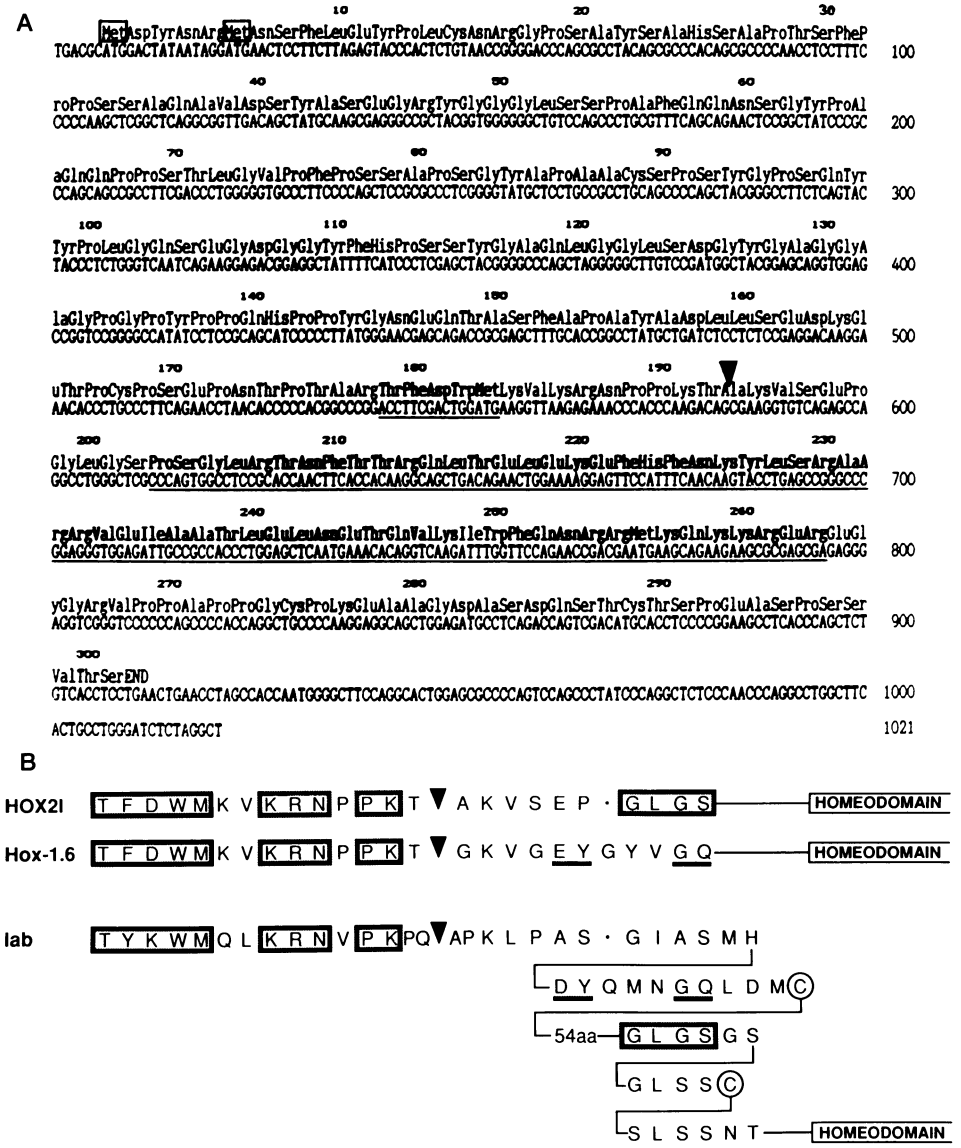


Fig. 8. A) A HOX2I cDNA clone. Symbols as in Fig. 6. Two possible initiating Met are boxed. B) Alignment of corresponding domains of HOX2I, the murine Hox-1.6 (25,28) and Drosophila lab (26,29) homeoproteins. Homologous peptides are framed or underlined. An arrowhead represents the splice site and two cysteine residues are circled. Dots indicate deletions. 54 aa indicates a stretch of 54 amino acid residues. Only one of several possible alignments of the lab domain downstream from the splice site is shown.

Fig. 9. Conserved domains (shaded boxes below the line) in homeoproteins. The amino-terminal domain may consist of 5-25 amino acid residues. The central domain extends from a few residues preceding the conserved YPWM pentapeptide to five residues following the HD.

3E, 4A and 4B. This pentapeptide is present in Drosophila homeotic genes but not in non-homeotic ANT-C genes containing an HD like fushi tarazu (ftz), zerknullt (zen) and bicoid (bcd). It appears that the pre-box, rather than the homeobox itself, discriminates true Drosophila homeotic genes (9).

This small domain might play a functional role complementary to, or in conjunction with, that played by the HD. The two domains are encoded in two different exons and are separated by a variable number of amino acid residues. The cooperation of the two domains may be independent of the distance, provided that this is not too large. Alternatively, different distances may be required for the specific function of various homeoproteins. Often the region between the YPWM homeopeptide and the HD is conserved in homeoproteins of the same group. We have shown that this is also the case for the group including HOX2I, Hox-1.6 and lab (Fig. 8B).

In summary, two domains are generally conserved in homeoproteins (33) and appear to be relevant in assigning homeoproteins to the various groups: the amino-terminal and a large domain spanning the homeopeptide, the HD and the downstream pentapeptide (Fig. 9).

Acknowledgments

We wish to thank Renato Somma for assistance with the computer. This work was supported by Progetto Finalizzato CNR "Biotecnologia e Biostrumentazione", the CNR Special Project "Human Genome", the Second AIDS Project of Ministero della Sanità and the Italian Association for Cancer Research AIRC. MDE, MP, EM and VN are recipients of an AIRC fellowship and DA is recipient of a fellowship from Fondazione Anna Villa Rusconi.

*To whom correspondence should be addressed

References

1. Levine, M. and Hoey, T. (1988) Cell 55, 537-540.
2. Akam, M. (1989) Cell 57, 347-349.
3. Beachy, P.A., Krasnow, M.A., Gavis, E.R. and Hogness, D.S. (1988) Cell 55, 1069-1081.
4. Hoey, T. and Levine, M. (1988) Nature 332, 858-861.
5. Herr, W., Strum, R.A., Clerc, R.G., Corcoran, L.M.,

Baltimore, D., Sharp, P.A., Ingraham, H.A., Rosenfeld, M.G., Binney, M., Ruvkun, G. and Horvitz, R.L. (1988) Genes Dev. 2, 1512-1515.
6. Robertson, M. (1988) Nature 336, 522-524.
7. Scott, M.P., Tamkun, J.W. and Hartzell, G.W. (1989) BBA Rev. Cancer 989, 25-48.
8. Gehring, W.J. and Hiromi, Y. (1986) Ann. Rev. Genet. 20, 147-173.
9. Boncinelli, E., Acampora, D., Pannese, M., D'Esposito, M., Somma, R., Gaudino, G., Stornaiuolo, A., Cafiero, M., Faiella, A. and Simeone, A. (1989) Genome 31, 728-743.
10. Poustka, A., Rackwitz, H.R., Frischauf, A.M., Hohn, B. and Lerach, H. (1984) Proc. Natl. Acad. Sci. USA 81, 4129-4133.
11. Simeone, A., Pannese, M., Acampora, D., D'Esposito, M. and Boncinelli, E. (1988) Nucl. Acids Res. 16, 5379-5389.
12. Mavilio, F., Simeone, A., Boncinelli, E. and Andrews, P.W. (1988) Differentiation 37, 73-79.
13. Simeone, A., Acampora, D., D'Esposito, M., Faiella, A., Pannese, M., Scotto, L., Montanucci, M., D'Alessandro, G., Mavilio, F. and Boncinelli, E. (1989) Molec. Reprod. Dev. 1, 107-115.
14. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).
15. Sanger, R., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
16. Acampora, D., Pannese, M., D'Esposito, M., Simeone, A. and Boncinelli, E. (1987) Hum. Reprod. 2, 407-414.
17. Boncinelli, E., Somma, R., Acampora, D., Pannese, M., D'Esposito, M., Faiella, A. and Simeone, A. (1988) Hum. Reprod. 3, 880-886.
18. Cannizzaro, L.A., Croce, C.M., Griffin, C.A., Simeone, A., Boncinelli, E. and Huebner, K. (1987) Am. J. Hum. Genet. 41, 1-15.
19. Oliver, G., Sidell, N., Fiske, W., Heinzmann, C., Mohandas, T., Sparkes, R.S. and De Robertis, E.M. (1989) Genes Dev. 3, 641-650.
20. Graham, A., Papalopulu, N. and Krumlauf, R. (1989) Cell 57, 367-378.
21. Duboule, D. and Dollè, P. (1989) EMBO J. 8, 1497-1509.
22. Hart, C.P., Fainsod, A. and Ruddle, F.H. (1987) Genomics 1, 182-195.
23. Mavilio, F., Simeone, A., Giampaolo, A., Faiella, A., Zappavigna, V., Acampora, D., Poiana, G., Russo, G., Peschle, C. and Boncinelli, E. (1986) Nature 324, 664-668.
24. Krumlauf, R., Holland, P.W., McVey, J.H. and Hogan, B.L.M. (1987) Development 99, 603-617.
25. La Rosa, G.J. and Gudas, L.J. (1988) Mol. Cell. Biol. 8, 3906-3917.
26. Mlodzik, M., Fjose, A. and Gehring, W.J. (1988) EMBO J. 7, 2569-2578.
27. Fritz, A.F., Cho, K.W.Y., Wright, C.V.E., Jegalian, B.G. and De Robertis, E.M. (1989) Dev. Biol. 131, 584-588.
28. Baron, A., Featherstone, M.S., Hill, R.E., Hall, A., Galliot, B. and Duboule, D. (1987) EMBO J. 6, 2977-2986.
29. Diederich, R.J., Merrill, V.K.L., Pultz, M.A. and Kaufman, T.C. (1989) Genes Dev. 3, 399-414.

30. Kongsuwan, K., Webb, E., Housiaux, P. and Adams, J.M. (1988) EMBO J. 7, 2131-2138.
31. Akam, M., Dawson, I. and Tear, G. (1988) Development 104 Supplement, 123-133.
32. Slack, J.M.W. (1983) From egg to embryo. Cambridge University Press, Cambridge.
33. Wright, C.V.E., Cho, K.W.Y., Oliver, G. and De Robertis, E.M. (1989). Trends Biochem. Sci. 14 (2), 52-56.
34. Fjose, A., Molven, A. and Eiken, H.G. (1988) Gene 62, 141-152.
35. Wedden, S.E., Pang, K. and Eichele, G. (1989) Development 105, 639-650.