Volume 336, number 1, 107–110 FEBS 13385 December 1993

The *HOX* complex neighbored by the *EVX* gene, as well as two other homeobox-containing genes, the *GBX*-class and the *EN*-class, are located on the same chromosomes 2 and 7 in humans

Toshikazu Matsui[a,b], Momoki Hirai[c], Masami Hirano[b], Yoshikazu Kurosawa[a,*]

[a]Institute for Comprehensive Medical Science and [b]*Department of Internal Medicine, Fujita Health University, Toyoake, Aichi 470-11, Japan*

[c]*Department of Anthropology, Graduate School of Science, The University of Tokyo, Bunkyo-ku, Tokyo 113, Japan*

Received 5 October 1993

Two newly identified human homeobox-containing genes, *GBX1* and *GBX2*, are closely related genes, as are members of the other homeobox genes, *EN-1* and *EN-2*. *GBX1* and *EN-2* have been mapped to chromosome 7q36. The present study shows that *GBX2 was* mapped to chromosome 2q37. *EN-1* was mapped to chromosome 2q14. Moreover, two *HOX* complexes neighbored by the *EVX* gene, *HOXA* and *HOXD*, are located at chromosome 7p15-p14 and 2q31-q37, respectively. Thus, it is possible that these homeobox genes were linked to each other on an ancestral genome and that the ancestral chromosome segment was duplicated during evolution.

Homeobox gene; Linkage group; Evolution

1. INTRODUCTION

The homeobox was first identified as a common sequence in several homeotic and segmental genes of *Drosophila*, such as *Antp*, *Ubx* and *Ftz* [1]. Around forty homeobox genes have been isolated from *Drosophila* to date [2]. Mouse and human homeobox genes homologous to those of *Drosophila* have also been isolated by cross-hybridization with *Drosophila* probes [3]. In addition, several homeobox genes, the homologues of which have not yet been isolated from *Drosophila*, have been identified in mouse and human by various methods, namely, hybridization with degenerate oligonucleotides that correspond to the helix 3 region of the homeodomain [4], the polymerase chain reaction (PCR) method [5], and a search for genes localized in the vicinity of the break point of chromosome translocation observed in leukemia [6]. The *Antp*-class genes form the largest family of homeobox genes in the human genome and they are clustered in four complexes, *HOXA*, *HOXB*, *HOXC* and *HOXD* [7]. The four *HOX* complexes appear to have been created by multiplication of a primordial gene cluster that corresponds to the *bithorax* complex and the *antennapedia* complex in *Drosophila* [8]. At the *HOM* and *HOX* loci, the clustering of genes and their order seem to be essential to development since these characteristics are retained in all vertebrates and arthropods. Two human homeobox genes, *EVX1* and *EVX2*, are also linked to the *HOXA* and *HOXD* clusters on chromosomes 7 and 2, respectively [9,10] In vertebrates, in addition to *HOX* genes, there are many examples of the duplication or multiplication of homeobox genes as follows: *EN-1* and *EN-2* [11], *Cdx1*, *Cdx2* and *Cdx3* [12]; *HOX7.1* and *HOX8.1* [13]; *Emx1* and *Emx2* [14]; *Gsh1* and *Gsh2* [14]; *Mox1* and *Mox2* [15]; *PBX1*, *PBX2* and *PBX3* [16]; *LFB1* and *LFB3* [17]; several *Pax* genes; and many *Oct* genes. These proliferated genes seem to be located randomly at various chromosomal loci, based on the present knowledge [18,19].

In a previous study [20], we identified two new homeobox genes, *GBX1* and *GBX2*, in the human genome by the PCR method. We characterized *GBX1* and mapped it to chromosome 7q36.1 [20]. The member of another class of homeobox genes, *EN-2*, has also been mapped to this region [21]. Since both the *EN*-class and the *GBX*-class of homeobox genes have two members, we suspected that *GBX2* might be mapped to a locus similar to that of *EN-1*, and that members of these two classes of homeobox genes could represent paralogous linkage groups.

2. MATERIALS AND METHODS

The *GBX2* probe was described in a previous paper [20]. A human genomic library was screened with the *GBX2* probe according to the Benton–Davis method [22]. Nucleotide sequence was determined by the dideoxy chain-termination method [23]. The details of the condi-

*Corresponding author. Fax: (81) (562) 93 8835.

Abbreviations: PCR, polymerase chain reaction; FISH, fluorescence in situ hybridization.

The sequence reported in this paper has been deposited in the EMBL data base (accession no. U02080).

tions for fluorescence in situ hybridization (FISH) are described elsewhere [24,25]. A fragment corresponding to human *EN-1* was amplified by PCR using two primers; 5'-GACAAGCGGCC(G/T)CG(G/C)AC(A/G)GC(G/C)TT(C/T)AC-3' and 5'-TGGTTGTACAG(T/G)-CCCTG(G/T)GCCATGAG-3'. A human genomic library was screened with this amplified fragment as a probe, and an *EN-1*-containing clone was isolated.

3. RESULTS

3.1. *Structure of the human GBX2 gene*

Preparation of the *GBX2* probe was described in a previous paper [20]. A human genomic library was screened with the *GBX2* probe. The nucleotide sequence of a homeobox-containing region was determined and indicated in Fig. 1a. Comparison of encoded amino acid sequences between *GBX1* and *GBX2* showed that the genes are closely related to one another (Fig. 1b) The degree of identity of amino acid residues in the homeobox, 58/60, is almost the same as that between *EN-1* and *EN-2* (57/60).

3.2. *Mapping of the human GBX2 and EN-1 genes*

The location of *GBX2* was determined by FISH. As shown in Fig. 2a, *GBX2* was mapped to chromosome 2q37. One of the HOX complexes, *HOX4D*, has been mapped to 2q31–37 [26]. In the mouse, this complex is located on chromosome 2 [27]. However, the mouse *En-1* gene is located on chromosome 1 [28], and the long arm of human chromosome 2 corresponds to mouse chromosomes 1 or 2, as indicated by comparative genome mapping [18,19]. In fact, human *EN-1* is located on chromosome 2 [21]. Very recently, *EN-1* was mapped to human chromosome 2q13-q21 using a mapping panel of rodent/human cell hybrids [29]. We examined the location of *EN-1* on human chromosome by FISH. As shown in Fig. 2b, *EN-1* was mapped to chromosome 2q14, that supported the above observation [29]. Thus, we concluded that both *GBX1* and *EN-2* are located at chromosome 7q36 while both *GBX2* and *EN-1* are located on the same chromosome but at different loci, 2q37 and 2q14, respectively.

4. DISCUSSION

In the mouse, several alleles in which mutations cause abnormalities in skeletal development are known to be located in the vicinity of either *En-1* or *En-2* [30]. In the vicinity of *En-1*, there exists *dominant hemimelia (Dh)*, a mutation that causes skeletal abnormalities in the hind limbs. In the vicinity of *En-2*, there exist two closely linked mutations, *hemimelic extratoes (Hx)* and *hammertoe (Hm)*, which cause skeletal defects in all four limbs. Martin et al. [30] reported that the distance between the *En-1* and *Dh* genes and that between the *En-2* and *Hx* genes are 0.28 cM and 1.1 cM, respectively, and they proposed that *En-1-Dh* and *En-2-Hx* represent paralogous linkage groups that evolved after duplication of a common ancestral chromosome segment [30]. When we started the present experiment, we suspected that the ancestral segment may have contained members of both the *EN*-class and the *GBX*-class of homeobox genes. However, *EN-1* and *GBX-2* are located on the same chromosome but not closely linked to each other. One of the genes, in which a mutation causes defects of the development of brain and face, holoprosencephaly, has been mapped to human chromosome 7q36 [31].

It has been argued that *HOX*-containing regions on chromosomes 2, 7, 12 and 17 arose by regional duplication or by tetraploidization and that genes for collagen are contained in these paralogous chromosomal regions [32,33]. The location of *GBX2* seems to be close to that of the *HOXD* complex. Although, in mouse, the *HoxD*

a)

```
 P  G  S  S  Q  G  G  R  P  G  P  R  G  G  G  D  P  A  E  Q  R  R  A
CCAGGCAGCTCACAAGGAGGAAGACCCGGGCCACGGGGTGGAGGAGACCCCGCCGAGCAGCGGCGCGCG
 G  S  T  T  S  T  G  K  N  R  R  R  R  T  A  F  T  S  E  Q  L  L  E
GGCAGCACCACGTCTACGGGCAAGAACCGGCGGCGGCGGACTGCCTTCACCAGCGAGCAGCTGCTGGAG
 L  E  K  E  F  H  C  K  K  Y  L  S  L  T  E  R  S  Q  I  A  H  A  L
CTAGAGAAGGAGTTCCACTGCAAAAAGTACCTCTCCTTGACCGAGCGCTCGCAGATCGCCCACGCCCTC
 K  L  S  E  V  Q  V  K  I  W  F  Q  N  R  R  A  K  W  K  R  V  K  A
AAACTCAGCGAGGTGCAGGTGAAAATCTGGTTCCAGAACCGACGGGCCAAGTGGAAACGGGTGAAGGCA
 G  N  A  N  S  K  T  G  E  P  S  R  N  P  K  I  V  V  P  I  P  V  H
GGCAATGCCAATTCCAAGACAGGGGAGCCCTCCCGGAACCCTAAGATCGTCGTCCCCATCCCTGTCCAC
 V  S  R  F  A  I  R  S  Q  H  Q  Q  L  E  Q  A  R  P  *
GTCAGCAGGTTCGCTATCAGAAGTCAGCATCAGCAGCTAGAACAGGCCCGGCCCTGAGGGG
```

b)

```
Antp RKRGRQTYTR YQTLELEKEF HFNRYLTRRR RIEIAHALCL TERQIKIWFQ NRRMKWKKEN
GBX1 SR-R-TAF-S E-L------- -CKK--SLTE -SQ-----K- S-V-V----- ---A---RIK
GBX2 NR-R-TAF-S E-L------- -CKK--SLTE -SQ-----K- S-V-V----- ---A---RVK
EN-1 D--P-TAF-A E-LQR-KA-- QA---I-EQ- -QTL-QE-S- N-S------- -K-A-I--AT
EN-2 D--P-TAF-A E-LQR-KA-- QT-----EQ- -QSL-QE-S- N-S------- -K-A-I--AT
```

Fig. 1. Nucleotide and amino acid sequences of the human *GBX2* gene. (a) Nucleotide and amino acid sequences of a homeobox-containing region of the *GBX2* gene. The homeobox region is boxed. (b) Comparison of amino acid sequences of the homeoboxes. The amino acid sequence of *Antp* is used as a reference [37]. Each bar indicates an identical amino acid to that in *Antp*. Amino acid sequences of *GBX1* and *EN-1/EN-2* are taken from refs. [20] and [34], respectively.

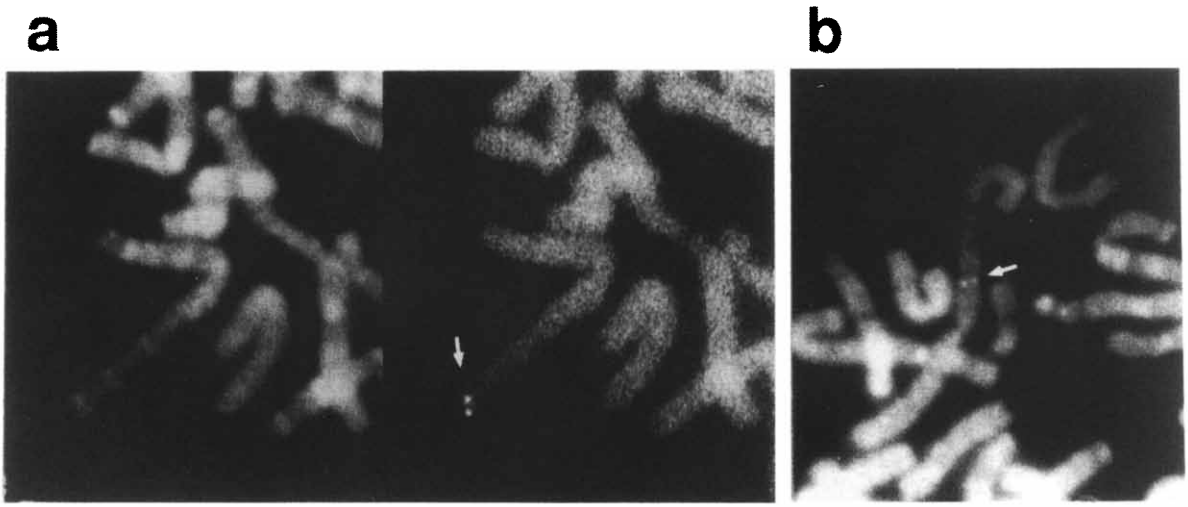

Fig. 2. Fluorescence in situ hybridization (FISH) of biotinylated human *GBX2* and *EN-1* probes on human chromosomes. (a) Mapping of *GBX2* to chromosome 2q37. A 2.5-kb *Kpn*I–*Acc*I fragment containing *GBX2* was used as the probe. Partial R-banded metaphase chromosome spread was viewed with a Nikon G-2A filter, showing R-bands (left). The same sample was viewed with a Nikon B-2E filter (right). An arrow indicates hybridization signals. (b) Mapping of *EN-1* to chromosome 2q14. An 11.3-kb *Bam*HI–*Eco*RI fragment containing *EN-1* was used as the probe for FISH. Partial metaphase spread was viewed with a Nikon B-2A filter, showing R-bands and hybridization signals (arrow).

complex and *En-1* are located on different chromosomes, the *Col6a-3* gene is mapped close to *En-1* [18,34]. Moreover, the *COL6A3* gene is located at human chromosome 2q37 [34]. The *EN-2/GBX1* cluster is located on the same chromosome as *HOXA* in human. Thus, it is possible that the *HOX* complex, the *EVX* gene, a gene for collagen, the *EN* gene and the *GBX* gene were linked to each other on an ancestral genome. Although members of multiplied homeobox genes other than *HOX* genes seemed to be located randomly at various chromosomal loci, some of them may have been organized similarly to *HOX*, *EVX*, *EN* and *GBX*. Increases in the complexity of the body and organs in vertebrates may have been achieved by such multiplication, with subsequent diversification, of regions of DNA that contained genes for regulatory and structural proteins [32]. Possible clustering of members of different classes of homeobox genes has been suggested for *Gsh*-3, *H6* and *HOX7* at human chromosome 4q16 [35]. Just before we submit this paper, Komuro and Izumo [36] reported that two newly identified homeobox genes, *Kbx* and *Imx*, are closely linked to each other.

Acknowledgements: We thank Dr. S. Ohno for valuable suggestions. We also thank Drs. Y. Takagi, I. Ishiguro and K. Fujita for their encouragement and Ms. A. Nagata for preparing the manuscript. This work was supported in part by grants from Japan Foundation for Aging and Health; Scientific Research on Priority Areas, Ministry of Education, Science and Culture of Japan; Fujita Health University (to Y.K.) and Special Coordination Funds for Promoting Science and Technology, Science and Technology Agency, Japan (to M.H.).

REFERENCES

[1] McGinnis, W., Levine, M.S., Hafen, E., Kuroiwa, A. and Gehring, W.J. (1984) Nature 308, 428–433.
[2] Scott, M.P., Tamkun, J.W. and Hartzell III, G.W. (1989) Biochem. Biophys. Acta 989, 25–48.
[3] Colberg-Poley, A.M., Voss, S.D., Chowdhury, K. and Gruss, P. (1985) Nature 314, 713–718.
[4] Singh, G., Kaur, S., Stock, J.L., Jenkins, N.A., Gilbert, D.J., Copeland, N.G. and Potter, S. (1991) Proc. Natl. Acad. Sci. USA 88, 10706–10710.
[5] Murtha, M.T., Leckman, J.F. and Ruddle, F.H. (1991) Proc. Natl. Acad. Sci. USA 88, 10711–10715.
[6] Kamps, M.P., Look, A.T. and Baltimore, D. (1991) Genes Dev. 5, 358–368.
[7] Scott, M.P. (1992) Cell 71, 551–553.
[8] Graham, A., Papalopulu, N. and Krumlauf, R. (1989) Cell 57, 367–378.
[9] Faielle, A., D'Esposito, M., Rambaldi, M., Acampora, D., Balsofiore, S., Stornaiuolo, A., Mallamaci, A., Migliaccio, E., Gulisano, M., Simeone, A. and Boncinelli, E. (1991) Nucleic Acids Res. 19, 6541–6545.
[10] D'Esposito, M., Morelli, F., Acampora, D., Migliaccio, E., Simeone, A. and Boncinelli, E. (1991) Genomics 10, 43–50.
[11] Joyner, A.L. and Martin, G.R. (1987) Genes Dev. 1, 29–38.
[12] German, M.S., Wang, J., Chadwick, R.B. and Rutter, W.J. (1992) Genes Dev. 6, 2165–2176.
[13] Monaghan, A.P., Davidson, D.R.,, Sime, C., Graham, E., Baldock, R., Bhattacharya, S.S. and Hill, R.E. (1991) Development 112, 1053–1061.
[14] Simeone, A., Gulisano, M., Acampora, D., Stornaiuolo, A., Rambaldi, M. and Boncinelli, E. (1992) EMBO J. 11, 2541–2550.
[15] Candia, A.F., Hu, J., Crosby, J., Lalley, P.A., Noden, D., Nadeau, J.H. and Wright, C.V.E. (1992) Development 116, 1123–1136.
[16] Monica, K., Galili, N., Nourse, J., Saltman, D. and Cleary, M.L. (1991) Mol. Cell. Biol. 11, 6149–6157.
[17] De Simone, V., De Magistris, L., Lazzaro, D., Gerstner, J., Monaci, P., Nicosia, A. and Cortese, R. (1991) EMBO J. 10, 1435–1443.
[18] Nadeau, J.H., Davisson, M.T., Doolittle, D.P., Grant, P. Hillyard, A.L., Kosowsky, M.R. and Roderick, T.H. (1992) Mammal. Genome 3, 480–536.
[19] O'Brien, S.J., Womack, J.E., Lyons, L.A., Moore, K.J., Jenkins, N.A. and Copeland, N.G. (1993) Nature Genet. 3, 103–112.

[20] Matsui, T., Hirai, M., Wakita, M., Hirano, M. and Kurosawa, Y. (1993) FEBS Lett. 322, 181–185.
[21] Logan, C., Willard, H.F., Rommens, J.M. and Joyner, A.L. (1989) Genomics 4, 206–209.
[22] Benton, D.W. and Davis, R.W. (1977) Science 196, 180–182.
[23] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[24] Takahashi, E., Hori, T., O'Connell, P., Leppert, M. and White, R. (1990) Hum. Genet. 86, 14–16.
[25] Ichihara, Y., Hirai, M. and Kurosawa, Y. (1992) Immunol. Lett. 33, 277–284.
[26] Cannizzaro, L.A., Croce, C.M., Griffin, C.A., Simeone, A., Boncinelli, E. and Huebner, K. (1987) Am. J. Hum. Genet. 41, 1–15.
[27] Stubbs, L., Huxley, C., Hogan, B., Evans, T., Fried, M., Duboule, D. and Lehrach, H. (1990) Genomics 6, 645–650.
[28] Hill, R.E., Jones, P.F., Rees, A.R., Sime, C.M. and Hastie, N.D. (1987) Cytogenet. Cell Genet. 44, 171–174.
[29] Köhler, A., Logan, C., Joyner, A.L. and Muenke, M. (1993) Genomics 15, 233–235.
[30] Martin, G.R., Richman, M., Reinsch, S., Nadeau, J.H. and Joyner, A. (1990) Genomics 6, 302–308.
[31] Gurrieri, F., Trask, B.J., van den Engh, G., Krauss, C.M., Schinzel, A., Pettenati, M.J., Schindler, D., Dietz-Band, J., Vergnaud, G., Scherer, S.W., Tsui, L.-C. and Muenke, M. (1993) Nature Genet. 3, 247–251.
[32] Ohno, S. (1970) in: Evolution by Gene Duplication, Springer, Berlin/Heidelberg/New York.
[33] Lundin, L.G. (1993) Genomics 16, 1–19.
[34] Schurr, E., Skamene, E., Morgan, K., Chu, M.-L. and Gros, P. (1990) Genomics 8, 477–486.
[35] Stadler, H.S., Padanilam, B.J., Buetow, K., Murray, J.C. and Solursh, M. (1992) Proc. Natl. Acad. Sci. USA 89, 11579–11583.
[36] Komuro, I. and Izumo, S. (1993) Proc. Natl. Acad. Sci. USA 90, 8145–8149.
[37] Laughon, A., Boulet, A.M., Bermingham, Jr. J.R., Laymon, R.A. and Scott, M.P. (1986) Mol. Cell. Biol. 6, 4676–4689.