Animal Cells and Systems 12: 261-267, 2008

Animal Cells
and
Systems

http://www.AnimalCells.or.kr

A Phylogenetic Analysis for Hox Linked Gene Families of Vertebrates

Sunwoo Kim, Gila Jung, Jaehyoun Lee, Hayoung Park, and Chang-Bae Kim*

Department of Life Science, Sangmyung University, Seoul 110-743, Korea

Abstract: The human chromosomes 2, 7, 12 and 17 show genomic homology around Hox gene clusters, is taken as evidence that these paralogous gene families might have arisen from a ancestral chromosomal segment through genome duplication events. We have examined protein data from vertebrate and invertebrate genomes to analyze the phylogenetic history of multi-gene families with three or more of their representatives linked to human Hox clusters. Topology comparison based upon statistical significance and information of chromosome location for these genes examined have revealed many of linked genes co-duplicated with Hox gene clusters. Most linked genes to Hox clusters share the same evolutionary history and are duplicated in concert with each other. We conclude that gene families linked to Hox clusters may be suggestion of ancient genome duplications.

Key words: molecular phylogeny, Hox cluster, vertebrates, evolution, genome duplication

It have generally been realized that the evolution of morphology of vertebrates is responsible for the modification of the genes and/or regulatory elements that control their development (Carroll et al., 2001). It have also been suggested that extensive gene/genome duplications at the origin of vertebrates results in widespread existence of gene families in extant vertebrates. In fact, in the invertebrates so far examined Hox genes of these are linked in a single cluster with some exceptions. But, by the emergence of higher vertebrates, the Hox gene cluster duplicated in either two rounds of genome duplication or third tetraploidization, resulting in the four or seven cluster state present in tetrapods and teleosts, respectively (Amores et al., 1998; Aparicio et al., 1997; McLysaght et al., 2002).

*To whom correspondence should be addressed.
Tel: +82-2-2287-2287; Fax: +82-2-2287-0700
E-mail: evodevo@smu.ac.kr

Mammalian Hox genes are homologous to clustered homeobox genes found in the *Antennapedia* and the *Bithorax* complexes of *Drosophila melanogaster* (Duboule and Dolle 1989; Graham et al., 1989). In both mouse and human, Hox genes are organized into four clusters (termed HOX A, B, C and D, each extending over 100 kb) that are located on four different chromosomes (Ruddle et al., 1994; Burglin 1994). Linked genes have been assigned up to 13 paralogous groups (as cognate groups 1-13) based on the fact that Hox genes show more sequence similarity to paralogous genes on other clusters than to genes linked on the same cluster. A hypothesis of mammalian Hox gene evolution suggested that the gene family arose through a series of gene duplications resulting in 13 paralogous groups linked in a cluster, followed by cluster duplication, which generated the current four-cluster state (Kappen et al., 1989; Schubert et a., 1993). Under this proposed model of evolution, cognate groups present on different clusters share a more recent common ancestor than genes within the same cluster, thus accounting for the higher level of sequence similarity observed among cognate groups on different clusters than among genes on the same cluster. The human chromosomes 2, 7, 12 and 17 show extensive intra-genomic homology, containing duplicate, triplicate and quadruplicate paralogous regions centered on the Hox gene clusters. These are taken as evidence that these paralogous gene families might have arisen from a single chromosomal segment through whole or block chromosome duplication events. The vertebrate genome likely has risen through a series of genome duplications (Ohno, 1970) which might be responsible for the increase in Hox cluster number from one to four (Holland et al., 1994; Ruddle et al., 1994). Reconstructing the history of the duplications and its relation to vertebrate evolution has been problematic due to the lack of informational sequence information from more extended adjacent gene families to Hox clusters. In this study, to see whether the four-fold paralogy seen on

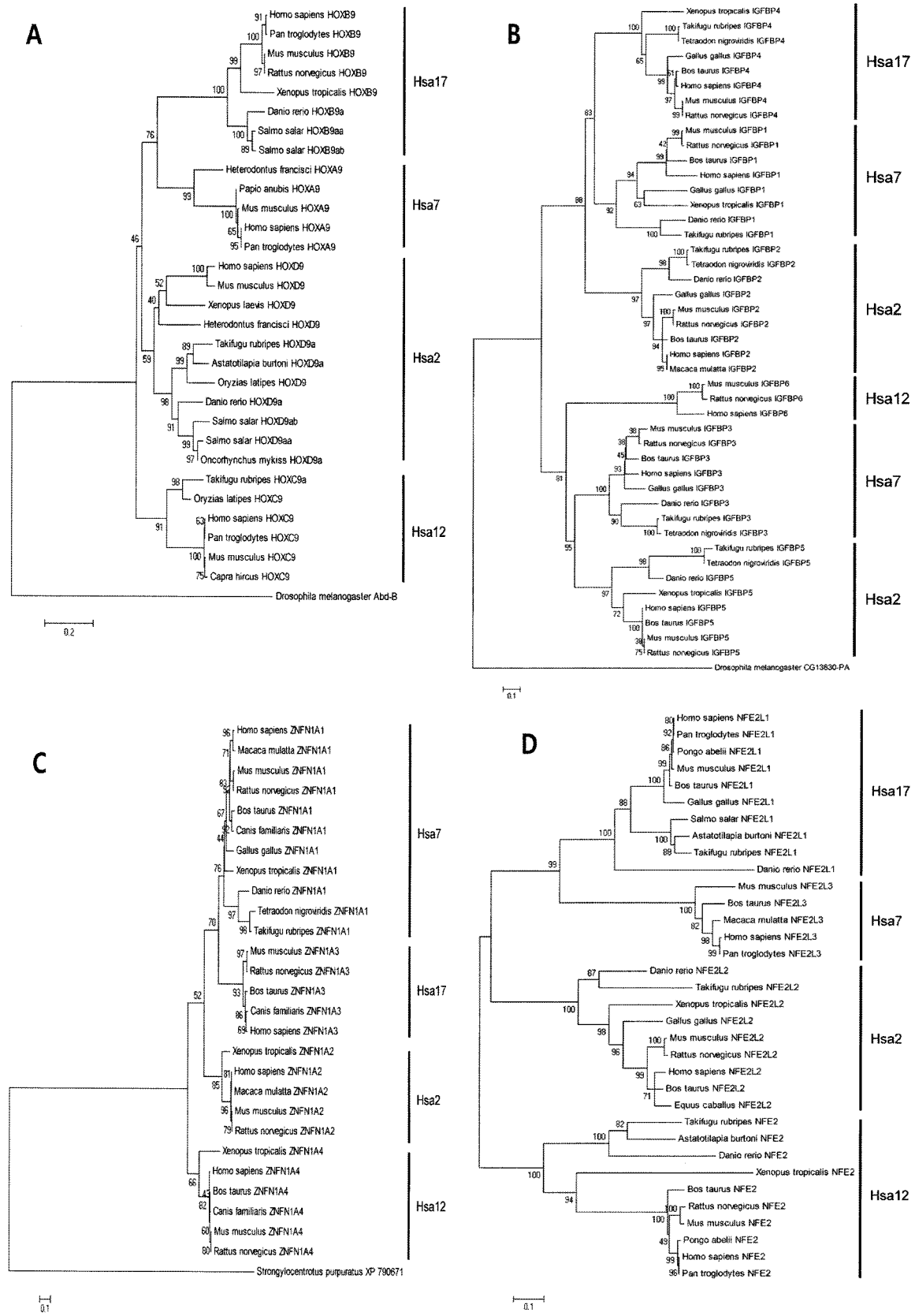

Fig. 1. Phylogenetic relationships of four gene families showing four members examined in this study. A, HOX9 gene; B, IGFBP; C, ZNFN1A; D, NFE2. Numbers on branches represent bootstrap values (based on 1,000 replications) supporting the branching patterns. Scale bar shows amino acid substitution per site.

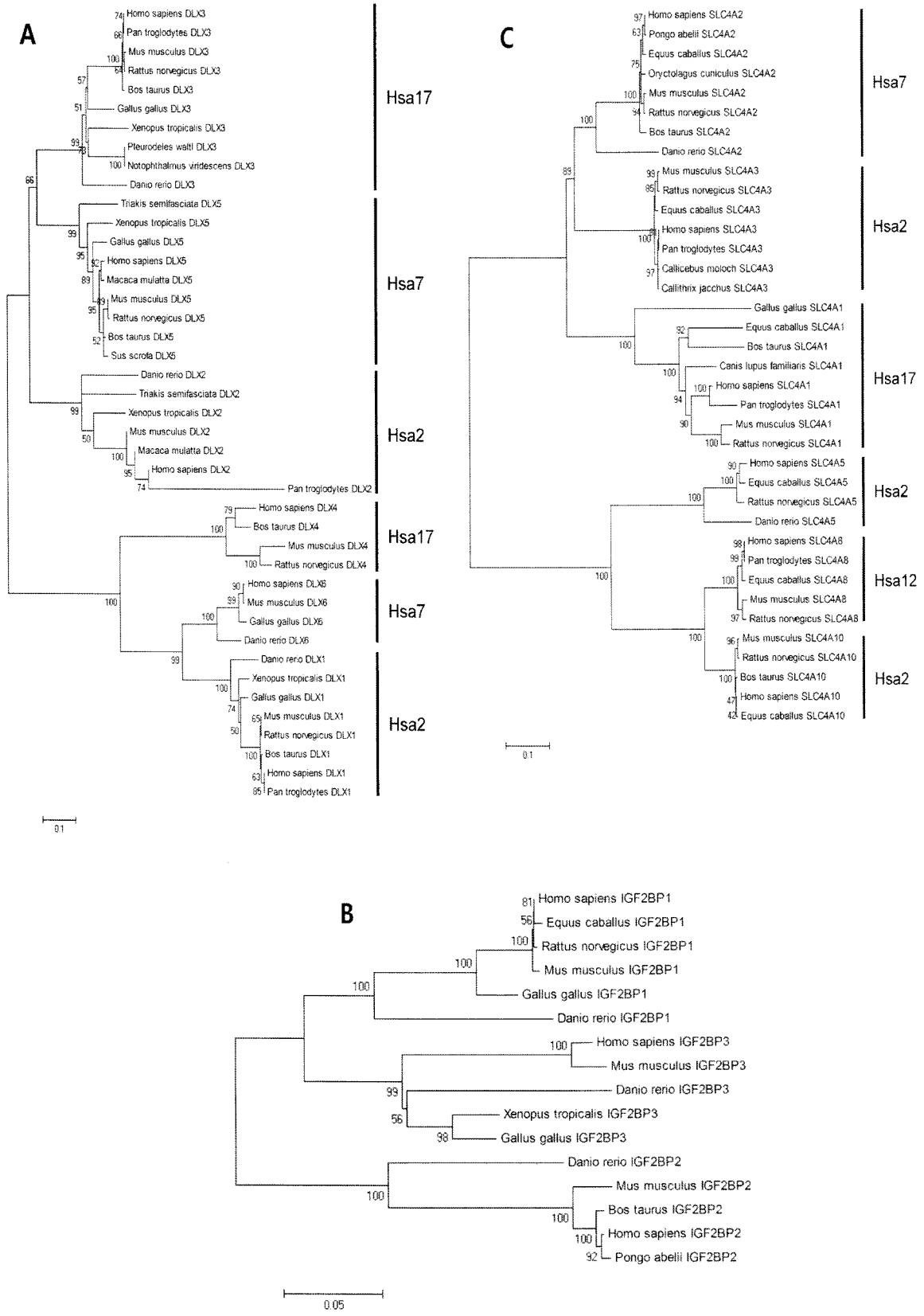

Fig. 2. Phylogenetic relationships of the gene families. A, DLX; B, IGF2BP; C, SLC4A.

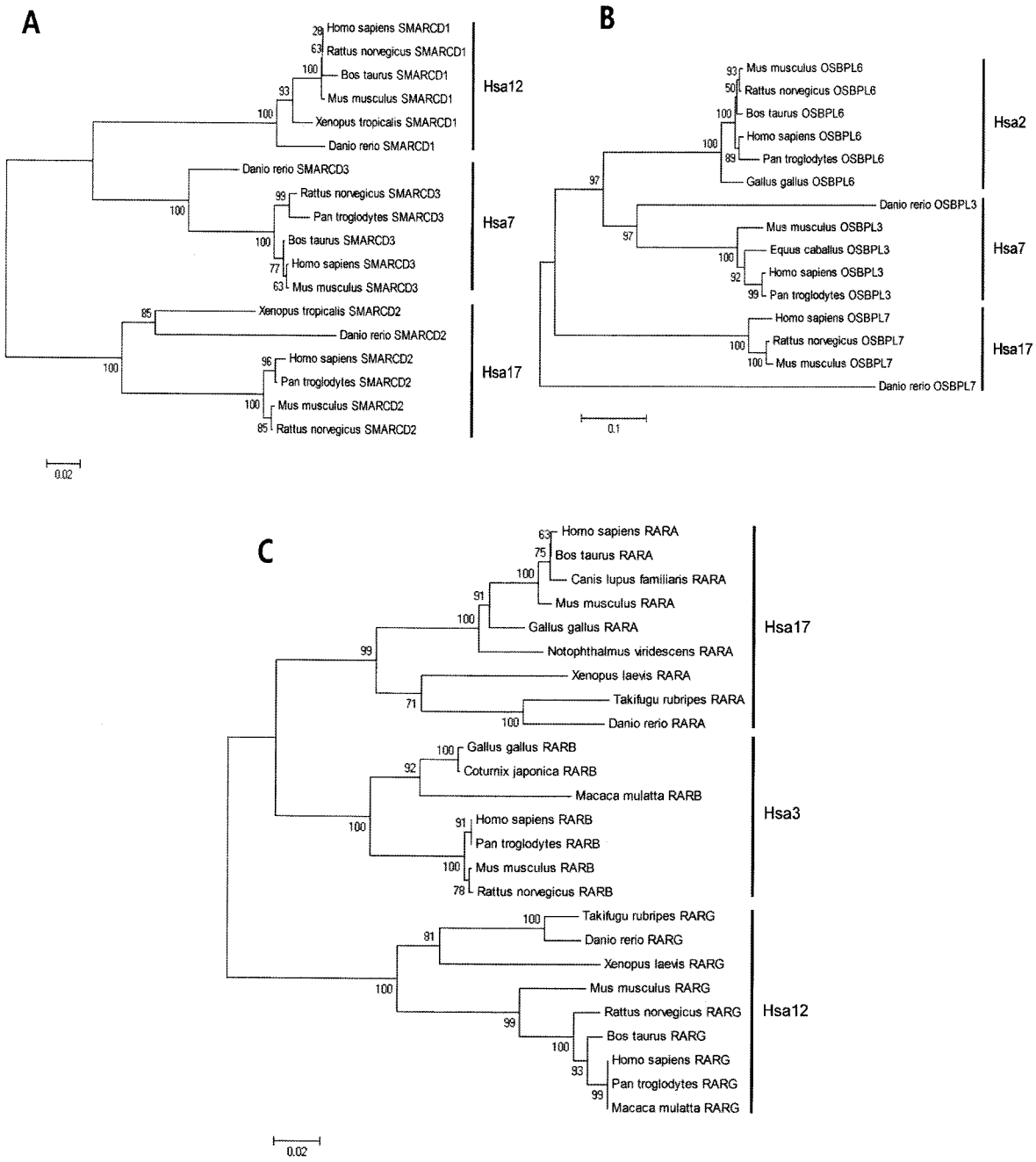

Fig. 3. Phylogenetic relationships of the gene families. A, SMARCD; B, OSBPL; C, RAR.

human Hox-bearing chromosomes is an outcome of genome duplications, we employed the phylogeny and positional information to know the consistencies among the phylogenies of 16 linked gene families including the Hox clusters and genome duplications.

MATERIALS AND METHODS

We have employed the currently available set of protein data for a wide variety of vertebrate and invertebrate genomes to analyze the phylogenetic history of 16 multi-gene families with three or more of their representatives linked to human Hox clusters. Protein sequences of human Hox linked gene families obtained from NCBI with BLAST and Ensembl database version 43. Protein domains in each sequence were identified by searches against the Pfam database and aligned using the Windows version of Clustal X 1.81. The phylogenetic relationships were reconstructed by Neighbor-Joining (NJ) method (Saitou and Nei, 1987) using MEGA program. JTT (Jones-Taylor-Thornton) matrix was used for calculation of genetics distance. Complete-deletion option was used. For statistical

Table 1. Summary of the analysis of gene families examined in this study

Gene Family Name	Chromosomal Position Information				Topology	Consistency with Genome Duplications
	Hsa2	Hsa7/3*	Hsa12	Hsa17		
HOX	HOXD9	HOXA9	HOXC9	HOXB9	(((7,17)2)12)	Yes
Collagen	COL3A1 COL5A2	COL1A2	COL2A1	COL1A1	(((12,17)7)2)	
DLX	DLX2 DLX1	DLX5 DLX6	- -	DLX3 DLX4	((7,17)2),((2,7)17)	Yes
ERBB	ERBB4	EGFR	ERBB3	ERBB2	(7,17),(2,12)	
GLI	GLI2	GLI3	GLI1	-	((2,7)12)	
HH	IHH	SHH	DHH	-	((2,7)12)	
IGF2BP	IGF2BP2	IGF2BP3	-	IGF2BP1	((7,17)2)	Yes
IGFBP	IGFBP2 IGFBP5	IGFBP1 IGFBP3	IGFBP6 -	IGFBP4 -	((7,17)2),((2,7)12)	Yes
INHB	INHBB	INHBA	INHBC INHBE	-	((2,7)12)	
ITGB	ITGB6	ITGB8/ITGB5*	ITGB7	ITGB3 ITGB4	((3,17)2)12)	
NFE2	NFE2L2	NFE2L3	NFE2	NFE2L1	(((7,17)2)12)	Yes
OSBPL	OSBPL6	OSBPL3	-	OSBPL7	((2,7)17)	Yes
RAR	-	RARB*	RARG	RARA	((3,17)12)	Yes
SLC4A	SLC4A3 SLC4A5 SLC4A10	SLC4A2	SLC4A8	SLC4A1	((2,7)17)	Yes
SMARCD	-	SMARCD3	SMARCD1	SMARCD2	((7,12)17)	Yes
SP	SP3	SP4/8	SP1	SP2	(((2,12)7)17)	
ZNFN1A	ZNFN1A2	ZNFN1A1	ZNFN1A4	ZNFN1A3	(((7,17)2)12)	Yes

Abbreviations for gene families: ERBB, ERBB receptor protein tyrosine kinase; GLI, GLI zinc-finger protein; HH, Hedgehog; IGF2BP, IGF-2 mRNA binding protein; IGFBP, Insulin like growth factor binding protein; INHB, Inhibin; ITGB, Integrin beta; NFE2, Nuclear factor erythroid 2; OSBPL, Oxystrol binding proteins; RAR, Retinoic acid receptors; SLC4A, Anion exchanger family; SMARCD, SWI/SNF-related, matrix-associated, actin-dependent regulators of chromatin; SP, Sp1 c2h2-type zinc-finger protein family; ZNFN1A, Zinc finger protein, subfamily 1A.

significance for branching pattern bootstrap was used based on 1,000 replications. The information of chromosomal location for gene families examined in this study obtained from the Sundstrom et al. (2008).

RESULTS AND DISCUSSSION

To perform independent testing of the hypothesis which supports four-fold paralogy regions in the human Hox-bearing chromosomes might be remnants of genome duplication, we conducted a phylogenetic analysis of gene families with representatives linked to three or four of the human Hox clusters. Gene families with paralogues linked to less than three Hox clusters have been left out because their occurrence is consistent with several alternative explanatory scenarios. Given the phylogenetic data, we sought to determine which genes could have duplicated simultaneously. For this test we reconstructed the phylogeny of vertebrate Hox clusters by the analysis of cognate group 9 (Fig. 1A). According to the resulting topology by neighbor-joining phylogenetic analyses with 16 gene families, we found that 12 families examined show a phylogeny that is concerted with the Hox cluster duplications based upon the Hox paralogous group 9 (Figs. 1-3; Table 1). Among these, 9 gene families are compatible with genome duplications with Hox.

The topology of Hox duplication shows the simultaneous duplication of members of four gene families, i.e. Hox, NFE2, and ZNFN1A, IGFBP which are represented with four cognate groups (Fig. 1B-D). Consistent with the compatibility in their tree topologies, each of the relevant genes is closely linked with the Hox cluster. This implies that the Hox linked genes share the similar evolutionary history as the Hox clusters and have arisen through the same duplication events that led to the Hox clusters. In the IGFBP the local duplication event could have occurred in early vertebrate evolution, while the gene pair was duplicated as a unit, most likely concomitantly with Hox. Two of IGFBP duplicates seem to have been lost before the separation of the branch leading to the mammals and the branch leading to the teleosts. This explanation is for the human genome with IGFBP gene pairs on chromosomes 2

and 7 and single genes on chromosomes 12 and 17.

The evolutionary pattern of six gene families with three members, DLX, IGF2BP, SLC4A, SMARCD, OSBPL and RAR is consistent with a quadruplication in early vertebrate evolution (Figs. 2, 3; Table 1). The phylogeny of the DLX well supports that DLX family underwent a local duplication before the chromosome duplications (Sumiyama et al., 2003). It can be established that DLX1, 4 and 6 form one series of paralogs and that DLX 2, 3 and 5 form another. Origination of IGF2BP by duplications before the vertebrate radiation has been suggested (Nielsen et al., 2002) and the present data supports the duplications. Even if SLC4A, SMARCD, OSBPL have three members in human chromosomes 2, 7 and 17, their trees support the genome duplications.

These gene families show that local duplications occur frequently and thereby complicate reconstruction of gene family histories that involve both local and chromosome duplications. Abbasi and Grzeschik (2007) proposed that gene families linked to Hox have duplicated independently with small chromosomal fragments and translocated to the present chromosomal locations. This suggestion is not compatible with the large scale researches dealing with large duplicated gene regions of vertebrate genomes (Dehal and Boore, 2005; Nakatani et al., 2007; Kuraku et al., 2009). Moreover the tree topology is further complicated by unequal evolutionary rates, gene conversion and loss of genes after duplication as shown in mammalian and teleost Hox genes (Powers and Amemiya, 2004; Hoegg and Meyer, 2005).

Because sequence based phylogenetic analyses and chromosome positions constitute two quite different types of information, they can be combined to deduce schemes in a more reliable way. The combined use of two types of information makes it possible to define old and new duplications, identify translocations of genes. Particularly gene families with very short sequences and highly conserved or rapidly diverging sequences, can be resolved by considering chromosomal position (Olinski et al., 2006; Larsson et al., 2008; Sundstrom et al., 2008).

We conclude that many more unrelated and expanded gene families to Hox gene clusters composing of the co-duplicated chromosomal regions may be suggestion of ancient genome duplications during chordate evolution. This supports expansion of two/three rounds of genome duplication harmonious to duplication of Hox gene clusters even if there have been many events related to gene losses obscuring duplication events.

ACKNOWLEDGMENTS

This research was supported by a grant from the Sangmyung University, 2008.

REFERENCES

Abbsi AA and Grzeschik KH (2007) An insight into the phylogenetic history of HOX linked gene families in vertebrates. *BMC Evol Biol* 7: 239.

Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL, Westerfield M, Ekker M, and Postlethwait JH (1998) Zebrafish hox clusters and vertebrate genome evolution. *Science* 282: 1711-1714.

Aparicio S, Hawker K, Cottage A, Mikawa Y, Zuo L, Venkatesh B, Chen E, Krumlauf R, and Brenner S (1997) Organization of the *Fugu rubripes* Hox clusters: evidence for continuing evolution of vertebrate Hox complexes. *Nat Genet* 16: 79-83.

Burglin TR (1994) A comprehensive classification of homeobox genes. In: Deboule D (ed), Guidebook to the Homeobox Genes, Oxford University Press, Oxford, pp. 25-72.

Carroll SB, Grenier JK, and Weatherbee SD (2001) From DNA to Diversity. Blackwell Science, Malden, MA. pp. 1-272.

Dehal P and Boore JL (2005) Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biology* 3: e314.

Duboule D and Dolle P (1989) The structural and functional organization of the murine HOX gene family resembles that of *Drosophila* homeotic genes. *EMBO J 8*: 1497-1505.

Graham A, Papalopulu N, and Krumlauf R (1989) The murine and *Drosophila* homeobox gene complexes have common features of organization and expression. *Cell* 57: 367-378.

Hoegg S and Meyer A (2005) Hox clusters as models for vertebrate genome duplication. *Trends Genet* 21: 421-424.

Holland PW, Garcia-Fernandez J, Williams NA, and Sidow (1994) A gene duplications and the origins of vertebrate development. *Development (Suppl.)*: 125-133.

Kappen C, Schughart K, and Ruddle FH (1989) Two steps in the evolution of antennapedia class vertebrate homeobox genes. *Proc Natl Acad Sci* USA 86: 5459-5463.

Kuraku S, Meyer A, and Kuratani S (2009) Timing of genome duplications relative to the origin of the vertebrate did cyclostomes diverge before or after? *Mol Biol Evol* 26: 47-59.

Larsson TA, Olsson F, Sundstrom G, Lundin LG, Brenner S, Venkatesh B, and Larhammar D (2008) Early vertebrate chromosome duplications and the neuropeptide Y receptor gene regions. *BMC Evol Biol* 8:184.

McLysaght A, Hokamp K, and Wolfe KH (2002) Extensive genomic duplication during early chordate evolution. *Nat Genet* 31: 200-204.

Nakatani Y, Takeda H, Kohara Y, and Morishita S (2007) Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome Res* 17: 1254-1265.

Nielsen FC, Nielsen J, Kristensen MA, Koch G, and Christiansen J (2002) Cytoplasmic trafficking of IGF- mRNA-binding protein by conserved KH domains. *J Cell Sci* 115: 2087-2097.

Ohno S (1970) Evolution by Gene Duplication. Springer Verlag, Heidelberg. pp. 1-160.

Olinski RP, Lundin LG, and Hallbook F (2006) Conserved synteny between the Ciona genome and human paralogons identifies large duplication events in the molecular evolution of the insulin-relaxin gene family *Mol Biol Evol* 23: 10-22.

Powers TP and Amemiya CT (2004) Evidence for Hox14 paralog group in vertebrates. *Curr Biol* 14: R183-184.

Ruddle FH, Bentley KL, Murtha MT, and Risch N (1994) Gene loss and gain in the evolution of the vertebrates. *Development (Suppl.)*: 155-161.

Saitou N and Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406-425.

Schubert FR, Nieseltstruwe K, and Gruss P (1993) The antennapedia-type homeobox genes have evolved from three precursors separated early in metazoan evolution. *Proc Natl Acad Sci USA* 90: 143-147.

Sumiyama K, Irvine SQ, and Ruddle F (2003) The role of gene duplication in the evolution and function of the vertebrate Dlx/distal-less bigene clusters. *J. Struct Funct Genomics* 3: 151-159.

Sundstrom G, Larsson TA, and Larhammar D (2008) Phylogenetic and chromosomal analyses of multiple gene families syntenic vertebrate Hox clusters. *BMC Evol Biol* 8: 254.

[Received November 3, 2008; accepted December 5, 2008]