Protein and Peptide Letters, Vol. 7, No. 2, pp. 123-131, 2000
Bentham Science Publishers B.V.
0929-8665/00 $ 15.00 + .00

DIVERGENCES IN HOMEODOMAINS OF HOX GENES AND VON BAER'S LAW

Rong-Qiao He* and Jun-Lan Yao

Lab of Visual Information Processing, Institute of Biophysics, The Chinese Academy of Sciences, Beijing 100101, China. Tel. 086 10 64889876, 086 10 64888531; Fax. 086 10 64877837; Email: herq@sun5.ibp.ac.cn

Key Words: Van Baer's law, Hox, homeodomain, development, evolution, homeobox, embryo, similarity, divergence.

Abstract: Divergences among homeodomains coded by Hox clusters of human beings have been analyzed with both methods of dynamic programming [1] and hydropathic mass [2]. Homeodomains of the anterior genes (Hox A-1 to -4), including the paralogs of Hox B and Hox D, showed more distinguishable divergences than those of the posterior genes (from Hox A-5 to -13). The homeodomains grew exponentially more divergent in progression from the paralogue group 1 to 13 along a Hox complex. The divergence of homeodomains of Hox B-1 between *C. elegans* and human was small, but it became larger between those of the posterior genes. It appears that Hox genes should be involved in Von Baer's law.

1. Introduction

Hox genes, organized in complexes [3-4], appear to control the organization of the embryo along the anterior-posterior axis in many animals [5], which are related to the *Drosophila* genes causing excitement in developmental Biology [6-7]. They are involved in the formation and development of some particular structures of animal bodies such as legs, wings, antennae, ribs, *etc*. Similarities of the organization and function among Hox genes of different animals have often been emphasized [8]. Hox genes hold evolutionary clues for us to understand the conservation of developmental mechanism. Although the relationship between Hox genes and chordate evolution has been studied in detail [9-10], few authors attempt to analyze similarities among homeo-domains from the anterior to posterior genes of a Hox cluster.

It has been demonstrated that homeobox genes [11-12], which are originally found in *Drosophila* and characterized by a highly conserved sequence of 180 nucleotides, are considered to play an important role in the genetic control of the development of *Drosophila* and other animals [13-14]. The homeodomain, consisting of 60 amino acids, confers the capacity on the proteins to bind specific DNA sequences and to regulate the expression of certain genes during embryogenesis [15-17]. The human and mouse Hox genes were mapped with four clusters, each containing approximately ten homeobox genes [18-20]. The order of Hox genes in the clusters corresponds to the order of expression along the body anterior-posterior axis [12,21]. The physical order of genes along the chromosome is collinear to their expression and function, spatially and temporally.

According to Von Baer's law [22], all developing vertebrates appear very similar shortly after gastrulation. It is only later in development that the special features of class, order, and finally species emerge. All vertebrate embryos have gill arches, notochords, spinal cords and pronephric kidneys, whose formation and development are regulated by Hox genes. Thus, Hox genes hold some important clues for us to investigate Von Baer's law. We have compared the divergences of

Hox homeodomains and observed that the divergences follow an exponential procedure from the anterior genes to the posterior ones and the anterior Hox genes of human beings share the highest similarities of the other species.

2. Materials and Methods

2.1. Analysis of similarities among homeodomains. Similarities of homeodomains of Hox are analyzed by two methods, one is Dynamic Programming method by Schwarty and Dayhoff [1], and the other is by Kyte and Doolittle [2], which are routinely used for estimation of similarities among biomolecules. The homeodomains of Hox (A, B, C and D) of human, mouse, *Xenopus*, *C. elegans* and *Drosophila* are used. All the peptide sequences employed are according to Rürglin [23]. Each homeodomain contains 60 amino acid residues without any sequence gaps. Thus, it is reasonable to use the methods to analyze similarities of the homeodomains of Hox genes.

2.2. Method of hydropathic mass. The 20 amino acids are alike in containing an α-carboxyl group and an α-amino group, but they are classified on the polarity of their side group [24]. Kyte and Doolittle used the hydropathic index of each amino acid and evaluated the hydrophilicity and hydrophobicity of a protein [2]. Although the method is not unique and embodies principles that have long been appreciated, its simplicity and graphic nature makes it a useful tool for the evaluation of protein structure. For analysis of similarity of peptides, hydropathic mass should be introduced, and it is regarded as the product of the mass of a side group and its hydropathic index:

$$\delta = hm \qquad (1)$$

where δ, h and m represent the hydropathic mass, the hydropathic index and the mass of the side chain of a residue, respectively (Table 1). According to Kyte and Doolittle [2], a triplet of the neighboring amino acid residues is used to calculate hydropathic mass along the sequence from N- to C-terminus, which is expressed as follows:

$$\eta = (h_n m_n + h_{n-1}\, m_{n-1} + h_{n-2}\, m_{n-2}) = (\delta n + \delta n\text{-}1 + \delta n\text{-}2) \quad (n \geq 1) \qquad (2)$$

where η and n represent value of triplet hydropathic mass and position of a residue from N- to C-terminus of the homeodomain. η depends on amino acid residue and its position. When a site mutation occurs at a position, it will cause a change in triplet hydropathic masses. However, it has been reported that not all site mutations in amino acid sequence lead to distinguishable changes in the conformation of peptide. Dayhoff *et al.* believes that there are some conservative amino acid groups, such as (i) L, I, M, V; (ii) G, S, P, T; (iii) F, Y, W; (iv) E, D, Q, N; (v) R, K, H, *etc.*, in which substitution of the amino acid residue each other would not influence the peptide conformation markedly [25].

Of the 20 amino acids, the hydropathic mass of Ile is the maximum, while that of Arg is the minimum. The difference between the two hydropathic masses is always the largest among the whole amino acids. Consequently, the difference of the triplet hydropathic masses between "-I-I-I-" and "-R-R-R-" is written as follows:

$$\Delta H = 3\,|M_I H_I - M_R H_R| = 3\,|\delta_I - \delta_R| \qquad (3)$$

The difference in the triplet hydropathic masses of amino acid residues is showed below:

$$\Delta\eta = |\delta_i + \delta_{i\text{-}1} + \delta_{i\text{-}2}| - |\delta_j + \delta_{j\text{-}1} + \delta_{j\text{-}2}| \qquad (4)$$

where i and j represent the position of amino acid residue. Hence, the divergence between the original site and the mutant can be evaluated by comparing the triplet hydropathic masses.

$$d = \frac{\Delta\eta}{\Delta H} \times 100\% \qquad (5)$$

where d represents the divergence between the two sites, and this divergence is concerned with polarity of the 3 residues. In comparison of the similarity or divergence among the homeodomains, the expression is as follows:

$$D = \left(\sum_{n=1}^{i} \Delta\eta\right)/(C \times \Delta H) \quad (i \geq 1)\ (C \geq 1) \qquad (6)$$

where D, i and C represent divergence between two peptides, the residue positions and numbers of amino acid residues, respectively.

2.3. Computer procedure for hydropathic mass. The sequences of homeodomains were analyzed by Dayhoff PAM-250 scale with Dolittle's well-established progressive alignment program, which was performed with SGI 4D/30 and SGI Indigo VAX/4500 computer.

Results and Discussion

3.1. Hydropathic masses of the homeodomains of human Hox clusters. As table 1 shows, the hydropathic mass of Ile is the highest and that of Arg is the lowest. However, that of Gly is approximately zero. Hydropathic mass is supposed to be a criterion to estimate the spatial position of a residue in a folded peptide or protein. Usually, a residue with a great negative hydropathic mass, especially with a negative triplet hydropathic mass, should have more opportunities to approach to the surface of a molecule. In contrast, the residue should be readily buried to the interior [2]. As we know, amino acid residues act mutually, especially the vicinal residues up- and down-stream in the sequence. The hydropathic masses of three residues are used as a group to calculate the polarity characterization of homeodomains although there are some remote residues in primary structure, which keep close to each other in spatial structure. How to estimate these interactions among the remote residues in the sequence needs further investigation.

Table 1. Some properties of amino acid side chains

Amino-acid	SCM*	H. index**	H. mass***
Gly	1	-0.4	~ 0
Ala	15	1.8	27
Val	43	4.2	180
Leu	57	3.8	217
Ile	57	4.5	257
Pro	42	-1.6	-58
Phe	91	2.8	255
Tyr	107	-1.3	-139
Trp	130	-0.9	-117
Ser	31	-0.8	-25
Thr	45	-0.7	-32
Cys	47	2.5	118
Met	75	1.9	143
Asn	58	-3.5	-203
Gln	72	-3.5	-252
Asp	59	-3.5	-207
Glu	73	-3.5	-256
Lys	72	-3.9	-281
Arg	100	-4.5	-450
His	81	-3.2	-260

* Data cited from *Biochemistry* by Lehninger A.L. *et al.* (1993) [24]. **Hydropathic index, Data provided by Kyte and Doolittle (1982) [2]. *** Hydropathic mass, on the basis of eq. 1.

The hydropathic masses of Hox homeodomains have a common characterization (Fig. 1), namely the domains contain some motifs with negative hydropathic mass. These motifs may be more hydrophilic and be exposed to the surface of the homeo-protein. In particular, the residues Ile-47 and Asn-51, which are the essential residues binding to the bases of TAAT of the target DNA, are located in the negative region of hydropathic mass with basic amino acid residues. This situation is supposed to be beneficial for the third helix of the homeodomain to combine with phosphate groups of DNA. In addition, around residue-30, there is another negative region of hydropathic masses with basic residues in the second α-helix of the homeodomain. In fact [26], when the homeodomain binds to the target DNA, the second helix approaches to the nucleotide strand where phosphate groups are located. It suggests that the two negative regions may support the essential residues to approach to the exterior and to combine readily with bases or phosphate groups of DNA.

As Figure 1 shows, the triplet hydropathic masses of residues 3-8, 28-32 and 47-55 are more conserved than those located in other regions of the homeodomains. Similar results were observed when we analyze the other homeodomains of Hox B, C and D, because the conserved residues are key amino acids. According to Kissinger [26], the motif (residues 47-55) inserts into DNA major groove, while the motif (residues 28-32) stabilizes the binding. Arg-3 and Arg-5 are key residues to bind the DNA minor groove. These residues are also located at the negative hydropathic mass region and this property improves them to insert into the grooves.

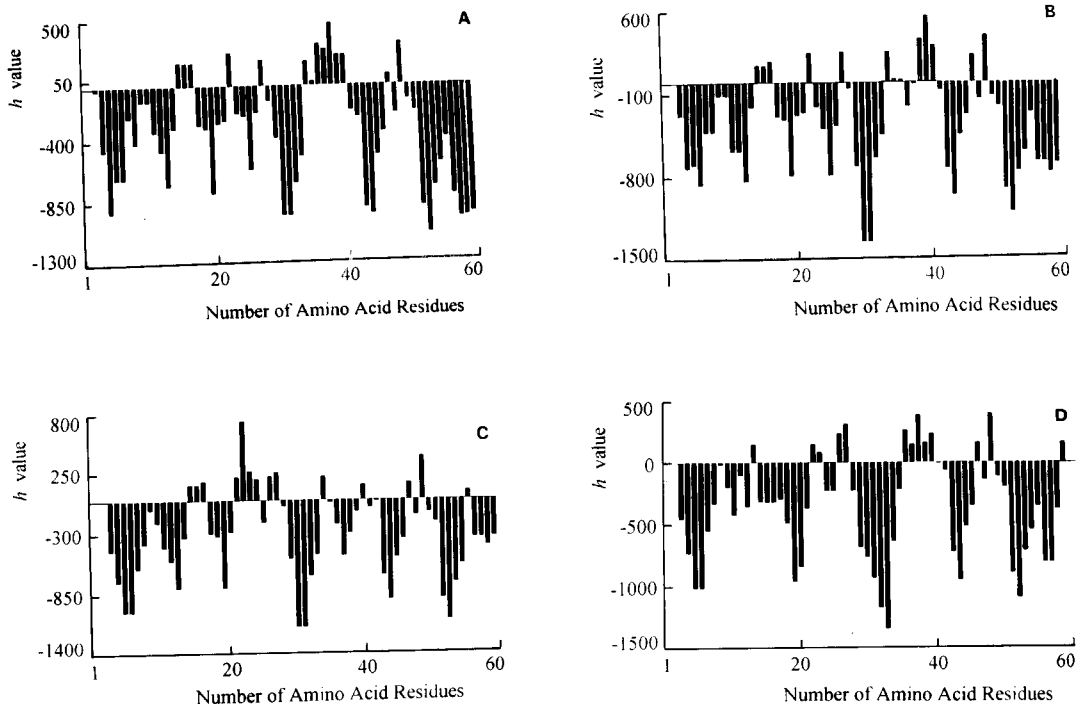

Fig. 1. Hydropathic masses of the homeodomains of human Hox A cluster.
On the basis of equation 2, hydropathic masses of Hox A homeodomains were calculated. The ordinate is the value of hydropathic masses and abscissa is position of amino acid residue. (A), (B), (C) and (D) represent the hydropathic masses of homeodomains of Hox A2, A5, A10 and A13, respectively.

However, we can see some divergences of hydropathic mass among the homeodomains of Hox A cluster, as well as the conserved regions. These differences may be one explanation that different homeobox molecules discriminate their target DNA. When comparing the similarities among the homeodomains from the anterior to posterior of the cluster, we can see the interesting changes in similarities of the domains along the developing temporal and spatial axis.

3.2. Similarities among homeodomains along Hox cluster. From the anterior to the posterior of Hox A, B and D clusters, the divergences among the homeodomains were analyzed by method of Schwarty and Dayhoff [1, 25]. The homeodomain divergences of the anterior genes (1-4) increase distinguishably, while those of the posterior genes (5-13) show a much less increase (Fig. 2). That is to say, the divergences among homeodomains increase much more greatly from Hox A1 to A4 (including the paralogs of Hox B and D) than those of the posterior domains in each cluster. However, the divergences among the homeodomains coded by Hox C cluster will follow a linear process when the domain of Hox C-4 is used as control (Fig. 2-C). At the same time, divergences among the genes in the clusters of Hox A, B and D also conform to linear processes when Hox A-4, B-4 and D-4 are used as controls, respectively. It suggests that Hox homeodomains evolve distinguishably from Hox 1 to 4 of the clusters, but gradually from Hox 5 to 13 (algebraic linear dependence on the position of Hox genes).

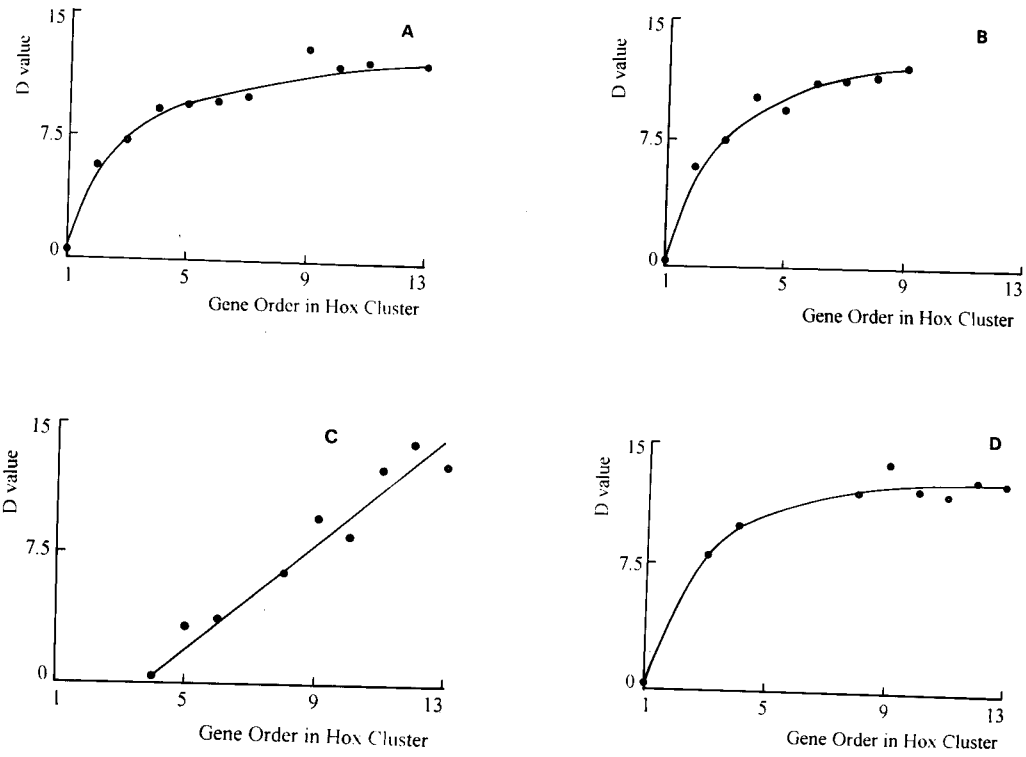

Fig. 2. Divergences among human Hox homeodomains by method of HM.

The homeodomains were referred as to Rürglin (1994). With equation 6, analysis of divergences of the homeodomains was carried out from the anterior to the posterior genes. For each cluster, Hox A1, B1 and D1 were respectively as controls. The ordinate is the genetic divergence of the domains and the abscissa is the gene position of Hox cluster. (A) Hox A; (B) Hox B; (C) Hox C, except Hox C4 was used as control. (D) Hox D.

Similar results were obtained by the method of hydropathic mass (Fig. 3). Divergences in polarity of the homeodomains along the clusters are coincident with those calculated by Dynamic Programming method. It appears that mutation frequency of homeobox molecules proceeds highly in the early evolution.

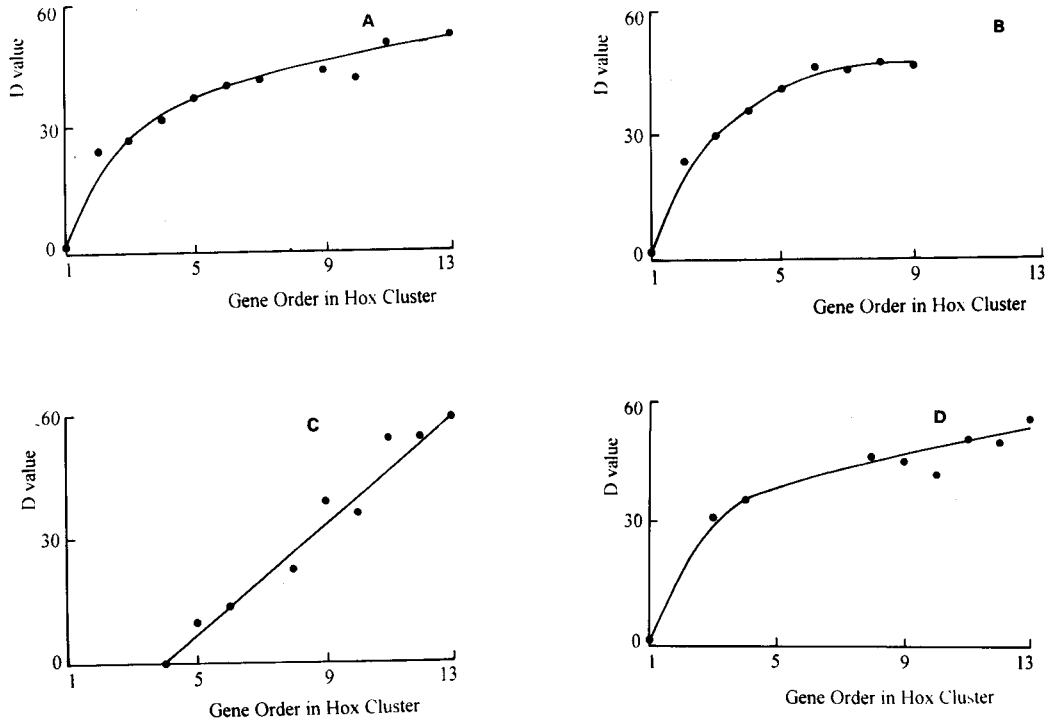

Fig. 3. Divergences among Hox homeodomains of human clusters by method of Dayhoff. Conditions were as for Fig. 2, except the dynamic programming method with Dayhoff PAM-250 scale (Schwarty 1978 and Dayhoff 1979) was used. The ordinate is the divergence values of the domains and the abscissa is the position of gene in a cluster. (A) Hox A; (B) Hox B; (C) Hox C, except Hox C4 was used as control. And (D) Hox D.

3.3. Earlier divergences of homeodomain conforming to an exponential process. As Table 2 shows, the numbers of different amino acid residues among Hox homeodomains increase along the clusters. However the site mutations occur mainly in the anterior genes. This view is supported by the results shown in Fig. 4. Divergences among homeodomains coded by the anterior genes of Hox A (from 2 to 4) follows an exponential process along the cluster. Following such an exponential procedure, the divergence of homeodomains between Hox A-1 and A-2 grows markedly. Similar results are obtained with the paralogs of Hox B and Hox D. Nevertheless, divergences among the homeodomains of Hox from A 5 to A13 (including Hox B 5 to B9) are getting smaller. As has been mentioned above, increase in the divergences follow an algebraic linear depend on the gene position. These also show that evolution of the homeodomains of Hox genes may

process slowly among the posterior genes in the clusters. According to the data shown in Fig. 2 and 3, for human and mouse, the divergences between homeodomains of Hox A-1 and A-2, between B-1 and B-2 as well as between D-1 and D-2 are approximately 6 (HM) and 25 (DH). And those between A-1 and A-9, between B-1 and B-9 as well as between

D-1 and D-9 are as high as 12 (HM) and 48 (DH), respectively. We can see that the divergences are getting larger while the genes are expressed along the Hox clusters during the early embryo development. Thus, we think that the molecular events of homeobox proteins with such a gradient divergence along the cluster, expressed spatially and temporally during embryo development, may be one important contributor to the heterogeneity of cells. On the basis of Von Baer's law[22],

Table 2. The numbers of different amino acid residues among human Hox homeodomains.

Hox*	A	B	C	D
1**	0	0	-	0
2	20	19	-	-
3	22	22	-	23
4	20	22	0	23
5	22	23	10	-
6	23	25	14	-
7	23	25	-	-
8		27	21	27
9	24	25	23	26
10	25	-	25	28
11	33	-	33	31
12	32	-	31	34
13	32	-	33	37

*For each cluster, The original gene was as control, for example Hox A1 was for Hox A cluster. ** The gene position of the cluster. All the sequences were refereed to Rürglin, (1994) [23].

early vertebrate embryos exhibit features common to the entire subphylum. As development progresses, embryos become recognizable as members of their class, their order, their family, and finally, their species. This is to say, development of an embryo is reflecting the evolution procedure of the species itself. But it takes much less time for the embryo to "repeat" the procedure during the short early development. Hence, it is reasonable that divergences of the anterior Hox genes conform to an exponential procedure, followed by a slow one in the posterior genes. So far, we might understand the significance of the exponential procedure that a developing embryo undergoes repeatedly a molecular "evolution" process along the temporal and spatial linearity with such an "instantaneousness", instead of millions of years for the species to take during evolution in nature. This implies that the ancient Hox molecules regulate the early embryo development. As we have seen that human embryo initially share the characteristics in common with fish and avian embryos, although the embryo never pass through a stage equivalent to an adult fish or bird [22].

3.4. Similarities of homeodomains among different species and Von Baer's law. Von Baer saw that different groups of animal shared certain common features during early embryonic development and that these features become more and more characteristic of the species as development proceeded. The divergences between the Hox homeodomains of *C. elegans* and those of mammals (Table 3) become more distinguishable from the anterior to posterior. The divergency value of Hox B-1 homeodomains between human and *C. elegans* is 4 (HM) and 21 (DH), however it becomes 10 (HM) and 58 (DH) of Hox B-9. Similar results are obtained with Hox B homeodomains between human and *Drosophila*. This is to say, while a mammalian embryo develops during the early period, the mammal will share the similar genetic events occurred in the round worm, but they depart from each other more and more remote following the process of embryonic development.

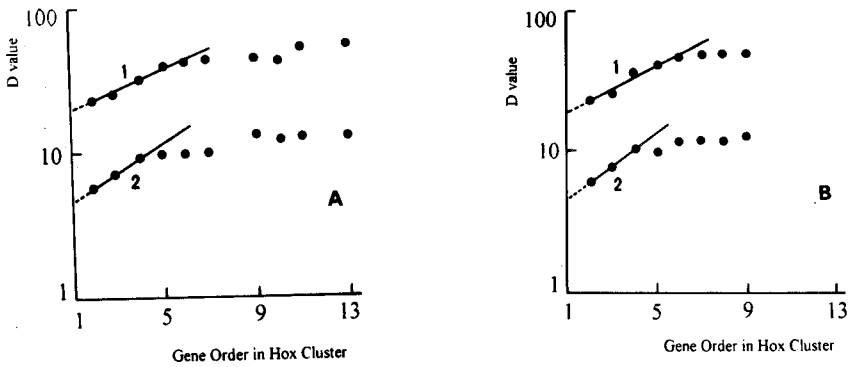

Fig. 4. Divergences among Hox homeodomains of human beings.

Conditions were as for Fig. 2, except the same data were analyzed in the semilogarithmic plot. The ordinate is the divergency value and the abscissa is the position of gene in a cluster. Curves 1 and 2 represent the data obtained by the dynamic programming method and the hydropathic mass method respectively. (A) Hox A and (B) Hox B.

Table 3. Similarities among Hox B homeodomains in different animals.

Hox	B1		B4		B7		B9	
	HM	DH	HM	DH	HM	DH	HM	DH
mouse	0.06	0.75	0.00	0.00	0.60	0.40	0.00	0.00
xenopus	-	-	1.60	4.97	1.29	1.46	0.18	0.96
Drosophila	3.21	12.15	1.12	5.79	0.93	0.66	6.11	24.87
C. elegans (round worm)	4.69	21.08	3.11	20.23	5.01	21.06	9.78	57.81

Methods of Dayhoff and hydropathic mass were employed. The human Hox genes were as controls. *Caenorhabditis elegans* was for the roundworm. Laevis of *Clawed frog* was for *Xenopus*. The genes of HOM cluster of *Drosophila* were used, which corresponded to B1, B4, B7 and B9 respectively. HM and DH represent the methods of hydropathic mass [2] and Dayhoff, respectively [1].

Here we would like to say Hox genes might be related to von Baer's principle. This view is supported by such observations. (1) The features, such as faces, gills, ribs and so on, shared by all vertebrate embryos during early development observed by Von Baer are strongly influenced by Hox genes. (2) The period after the gastrulation, during

which the embryos share the features, is almost the same as Hox genes are expressed, regulating the embryogenesis. And (3) the expression of the genes is parallel to the regularity of both spatial and temporal collinearities *in vivo* [4,8]. This corresponds to the increase of the heterogeneity of the embryo along the collinarities during the early embryonic development. The resemble features shared by all the embryos occur along the collinearities. It suggests that the early embryo development of vertebrate passes through such a procedure that the homeobox genes, which function in morphogenesis, are expressed from the ancient molecules to the current.

Acknowledgments: This project is supported by the Foundation of the Chinese Academy of Sciences, 95 Key Foundation and the Chinese Natural Science Foundation No. 39770254.

References

1. Schwarty R.M. and Dayhoff M.O. (1978) *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington D.C.), vol. **5**, Supp.3, pp.353-358.
2. Kyte J. and Doolittle R.F. (1982) *J. Mol. Biol.* **157**, 105-132.
3. Rabin M. and Hart C.P. (1985) *Nature* **314**, 175-178.
4. McGinnis W and Kuziora M. (1994) *Scientific American* **270**, 58-70.
5. McGinnis W., Levine M., Hafen E., Kuroiwa A and Gehring W.J. (1984a) *Nature* **308**, 428-433.
6. Sharkey M, Graba Y. and Scott M.P. (1997) Trends Genet., **13**, 145-151.
7. Holland P.W.H. and Garcia-Fernandez J. (1996) Dev. Biol. **173**, 382-395.
8. McGinnis W. and Krumlauf R. (1992) *Cell* **68**, 283-302.
9. Burke A.C. Nelson C. E., Morgan B.A. and Tabin C. (1995) Development, **121**, 333-346.
10. Holland P.W.H. and Garcia-Fernandez J. (1994) Development, 1994 Supplement: 125-133.
11. Gehring W.J. (1994) A history of the homeobox, in *Guidebook to the Homeobox Genes* 1st Ed by Duboule D. Oxford Univ. Press, Oxford, pp3-10.
12. Lewis E.B. (1978) *Nature* 276, 565-570.
13. McGinnis W., Garber R.L., Wirz J., Kuroiwa A. and Gehring W.J. (1984b) *Cell* **37**, 403-408.
14. Scott M. P. and Weiner A.J. (1984) *Proc. Natl. Acad. USA* **81**, 4115-4119.
15. Laughon A and Scott M.P. (1984) *Nature* **310**, 25-31.
16. Hiromi Y. and Gehring W.J. (1987) *Cell* **50**, 963-974.
17. Jaynes J. B. and O'Farrell P.H. (1988) *Nature* **336**, 744-749.
18. Balling R., Mutter G., Gruss P. and Kessel M. (1989) *Cell*, **58**, 337-347.
19. Bürglin T.R., Ruvkun G., Coulson A., Hawkins N.C., McGhee J.D., Schaller D., Wittman C., Müller F. and Waterston R.H. (1991) *Nature* **351**, 703-707.
20. Kenyon C. and Wang B. (1991) *Science* **253**, 516-517.
21. Krumlauf R. (1994) *Cell* **78**, 191-201.
22. Gilbert S.F. (1991) Early vertebrate development Neurulation and ectoderm, in *Develop-mental Biology*, 3rd Ed, Sinauer Associates, Inc. Massachusetts, pp. 155-157.
23. Rürglin T.R. (1994) A comprehensive classification of homeobox genes, in Guidebook to the homeobox genes, Ed by Duboule, D. Oxford Univ. Press, Oxford, pp.25-73.
24. Lelinger A.L., Nelson D.L. and Cox M.M. (1993) *Principles of Biochemistry* 2nd Ed, Worth Publishers, New York, pp.130.
25. Dayhoff M.O., Schwartz R.M. and Orcutt B.C. (1979) *Atlas of Protein Sequence and Structure*. ed. Dayhoff, M. O. (Natl. Biomed. Res. Found., Washington). pp.345-352.
26. Kissinger C.R., Liu B., Martin-Blanco E., Kornberg T.B. and Pabo C.O. (1990) *Cell* **63**, 579-590.

Received on February 1, 2000, accepted on April 7, 2000