**HOX-PRO DB: THE WAYS OF EVOLUTION OF ENSEMBLES OF HOMEOBOX GENES-CONTROLLERS OF DEVELOPMENT.**

SPIROV A.V.

The Sechenov Institute of Evolutionary Physiology and Biochemistry, 44 Thorez Prospect, St.Petersburg, 194223, Russia; spirov@iephb.ru;

Keywords: evolution, homeobox, genes, development, database, sea urins, caenorhabditis. elegans, Drosophila melanogaster, vertebrates, HOX-clusters, hox-networks

Nowadays homeobox genes were found for species of main invertebrate and vertebrate taxa. It is established that these genes play key role in time and space orchestration of genome expression during development. Auto- and crossregulatory functional interactions join homeobox genes into genetic networks. For the purpose to orange all available data on structure, functions, phylogeny and evolution of hox-genes, HOX-clusters and hox-networks we develop specialized database HOX-Pro. The HOX-Pro is aimed at analysis and classification of regulatory and coding regions in diverse homeobox genes-controllers of invertebrate and vertebrate development (1); comparative analysis of organization of "hox-based" genetic networks for *Caenorhabditis.elegans*, sea urchins, *Drosophila melanogaster* and vertebrates (2); analysis of phylogeny and evolution of homeobox genes and clusters (3).

Its main location is <http://www.iephb.ru/~spirov/hox_pro/hox-pro00.html>. The DB is also mirrored at <http://www.mssm.edu/molbio/hoxpro/new/hox-pro00.html>.

**1. Introduction**

Mutations within homeotic hox-genes in *Drosophila melanogaster* transform defined segments of the body into the character of an adjacent segments. Homologous genes there are in vertebrates, where it has been possible to mutate them, shifts in morphological character have also been produced.

The discovery of the first HOX cluster outside *Drosophila* causes great hopes. It was proposed that the homeobox would become a "Rosetta stone" for the study of animal development. This could enable us to read the epigenetic code of other animals on the basis of our understanding of *Drosophila* [1]. It is possible that this HOX-based epigenetic code is very ancient and was in place in the common ancestor of all modern animals. According to Slack, Holland and Graham, this character should be adopted as the defining character, or synapomorphy, of the kingdom *Animalia*.

We develop the HOX-Pro database for description of homeobox-genes controlling embryogenesis [2]. It contains a broad spectrum of information including pictures, schemes, movies and applets. Graphical representation of HOX-clusters and hox-based networks is accomplished in the form of flow diagrams, as well as in the form of the Java applets, that permits to emphasize the interacting genes in the network, to reflect the mode of gene actions, etc.

The distinctive feature of the HOX-Pro is its ability to serve not only as an information depository but also as tool for derivation of new knowledge by means of computer analysis. This communication is intended to show perspectives of comparative evolutionary approach to study functions and evolution of homeobox genes-controllers of development as it is presented in the HOX-Pro.

**2. HOX-clusters and hox-networks evolution**

No matter how they initiate development, as these embryos establish their body, the conserved "rainbow-like" expression pattern of the HOX-cluster genes begin to appear. However, despite wonderful conservatism of such pattern, HOX-clusters undergo essential rearrangements in evolution of main taxa. One can distinguish cluster and gene duplications, juxtapositions of duplicated portion of a gene and possible addition of *cis-*elements by transposons. Below we summarized these main ways of HOX-clusters and networks reorganization.

**2.1. HOX-clusters duplications**

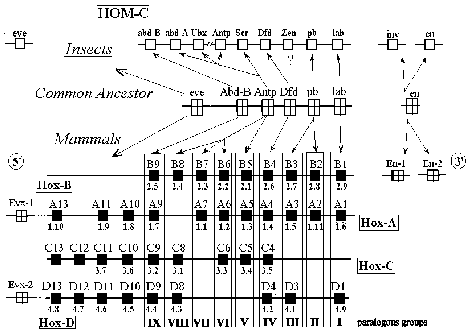
Four HOX clusters in mammals arose presumably by two or more duplications of one ancestral cluster (fig.1). By the conserved gene organization in each cluster and criteria of sequence conservation, it is possible to identify closely related groups of genes on different clusters - so called paralogous groups. Nine such groups (paralogous groups 1 through 9), are thought to be represented in the ancestral cluster. Nine of assumed 36 genes in these paralogous groups are missing.

In the teleost fish Fundulus heteroclitus, a total of 22 members from eight paralogous groups were found, consistent with as many as ten of 32 genes missing [3]. It demonstrates significant differences of the Fundulus clusters compared to those of mouse and humans. In the zebrafish there are four HOX clusters with as many as ten missing genes in paralogous groups 1 through 9 [4].

Within the HOX-cluster of Arthropods homeotic genes the two *Anip-*like genes which lack the function of the genes-selectors could be pointed out. These genes are *fushi tarazu* and *zerknult*(fig.1). These genes are used to be characterized as rapidly evolving as opposed to the other genes from the same cluster.

In connection with this, another rapidly evolving hox-gene, *even-skipped (eve)* should be mentioned. This gene controls the process of neurogenesis in arthropods. Beyond this, supposedly, ancient function of this gene, *eve* appears to be involved in the processes of segmentation of the early embryo of the *Diptera*insects. However, *eve* gene of *Diptera* is not linked with the HOX-cluster, whereas its homologues of vertebrates are localized in the *HoxA* and *HoxD* clusters (fig.1) as the 5' end genes. Possibly, during the evolution in the ancestors of arthropod *eve* lacked the physical and functional relations with the homeotic cluster, while at the evolution of vertebrates, such event has not take place. That is why, the rapidly evolving genes of the HOX/HOM-cluster are of particular interest from the positions of the molecular evolution.

Figure 1. Possible evolutionary relationships between insect and vertebrate clusters of homeobox genes. Alignment of the four vertebrate HOX clusters with the *Drosophila* HOM-C homeotic complex is presented. The letters above the boxes represent modern nomenclature. Below the boxes are the former mouse gene names.



**2.2. Juxtaposition duplicated portion of a gene to new cis-regulatory elements**

During duplication duplicated portion of a gene, including its coding region, is assumed to be juxtaposed to new *cis-*regulatory elements which change its expression and thus give it a new function - ultimately the activation of different set of genes [5].

The three paired-box and homeobox genes *paired*, *gooseberry* and *gooseberry neuro* have distinct developmental functions in *Drosophila* embryogenesis. However, despite the functional difference and the considerably diverged coding sequence of these genes, their proteins have conserved the same function [5]. The finding that the essential difference between genes may reside in their *cis-*regulatory regions exemplifies an important evolutionary mechanism of how function diversifies after gene duplication.

Other three clustered *Drosophila* genes homeotic *spalt*, *spalt adjacent*, *spalt related*, and their present *cis-*regulatory regions arose through a chromosomal rearrangement involving local duplication and transposition events in the 32F/33A region on the left arm of the second chromosome [6].

**2.3 Addition and distribution of control elements by transposons**

Transposable elements (TEs) - transposons and retroposons, are a major source of genetic change, including the creation of novel genes, the alteration of gene expression in development, and the genesis of major genomic rearrangements [7].

TE insertions can affect the expression patterns of endogenous genes by adding and distributing specific control elements throughout the host genome. Transposition of a TE into or near a particular host gene - possibly followed by an excision event leaving behind the TE's regulatory sequences - might impose novel developmental control on this host gene. Recent observations have shown that a number of sequences known mobile elements were frequently inserted into the DNA of a gene regions and can influence the regulation of a gene's expression [8].

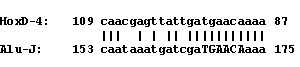
There are above twenty examples from *Drosophila*, sea urchin, human and mouse genomes that meet the following criteria: the element was inserted far in the past and thus the event is not a transient mutation (1); the element is a member of a large group of similar sequences (2); the inserted element now serves a useful function (3). All these cases are summarised in electronic table at <http://www.iephb.ru/~spirov/hox_pro/table.html>.

The four known *Spec* genes of sea urchin *Strongylocentrotus purpuratus* have "RSR" repeated sequences in their 5' regions [8]. The "*RSR*" element of the *Spec2a* gene has been examined carefully and includes an enhancer. The *RSR* enhancer is required for the tissue-specific expression of the *Spec2a* gene and contains four *Otx* binding sites for a *bicoid-*class homeodomain protein of sea urchins (*SpOtx*).

Two *Intracisternal A-Particle* (*IAP*, transposable element, the family of intracisternal A particles) insertions have been identified in WEHI-3B myeloid leukemic cells, one at the *interleukin-3* gene and another at the homeobox gene *HoxB-8 (Hox-2.4)* [9]. The provirus has inserted within the first exon of the gene and generated a *Hox-2.4* mRNA with a 5' sequence derived from the *IAP* long terminal repeat. Both proviral insertions have resulted in transcriptional activation of the adjacent genes, which appears to be a significant step in the leukemogenic process in these leukemic cells [9].

Vansant and Reynolds [10] have observed that *Alu* sequences include functional binding sites for retinoic acid (RA) receptors. The question is risen about the possible function thousands of RA receptor-binding sites in *Alu* sequences scattered throughout the human genome. On the other hand, there are well documented examples of mammalian/human genes that include in their promoters *Alu* sequences carrying the RA receptor- and estrogen receptor-binding sites (Britten, 1997).

To date, the RA-response elements (RAREs) identified in the vertebrate HOX-clusters are RARE of *HoxA-1* (murine and human), RAREs of the murine and human *HoxB-1* genes and RAREs conserved in the promoters of the human and murine *HoxD-4* genes. All these data is summarised at <http://www.iephb.ru/~spirov/hox_pro/hox-rares.html>. Conserved cluster from three diverged DR5 in the promoter region of human/murine *HoxD-4* gene shows sequence homology as compared with *Alu-J* and *K18 Alu* (See <http://www.iephb.ru/~spirov/hox_pro/alu2hoxd.html>). Particularly search of *Alu* repeats from REPBASE gives such high-scoring segment pairs:



This is the sequence from RARE1 box of *HoxD-4* [11]; "TGTTCA/TGAACA" is the first direct repeat.

**3. Discussion and Conclusions**

Even the cursory inspection of the materials on the functional organization of the hox-genes and their clusters can lead to some conclusions about mechanisms of their evolution for the main invertebrate and vertebrate groups. The conservatism of the structural and functional organization of the HOX-clusters even during the comparison of very diverged species is well known and has been repeatedly discussed in literature (See [2]). Hence, the revealed ways and mechanisms of evolutionary variability of the hox-genes and their ensembles are of fundamental interest.

In this connection, the fact of great attention is that the multiple massive duplications of the large regions of the chromosomes, including the HOX-clusters was accompanied not only by duplication, but also by the essential lack of the hox-genes. As this takes place, some of the rapidly evolving genes can physically and functionally move to the other loci and to the other genetic networks (as it possibly took place in the case of *eve* and *Hoxb-3* paralogous genes in the invertebrate evolution). During the evolution from the lower insects to the higher *Diptera*, *eve* gene, supposedly, returned to the segmentation genetic network, but remained outside the HOX-cluster (unlike the two paralogous mammalian genes *Evx-1* and *Evx-2* which are adjacent to the HOX-clusters).

The evolutionary instability of the HOM-clusters of arthropods and HOX-clusters of vertebrates could be evidently proved by the evolutionary recent insertions (such as the insertion of the fragment of the DNA, encoding the polypeptide similar to ?-subunit of the human ATP-ase into the mouse *Hox-2.7*gene [12]).

It is well known that the hox-ensembles of vertebrates differ from the same of invertebrates by the key role of the RA in the control of the hox-genes expression. On the other hand, the *Alu* sequences, represented in the vertebrate genomes by many thousands of copies, are sometimes carrying the functional RA-dependent enhancers. This fact together with the similarity in sequences of known RA-dependent enhancers in the mammalian hox-genes with some human *Alu* sequences, pointed out in the HOX-Pro, leads to assumption about the possible role of the *Alu* elements in the multiplication and distribution of the RA-response elements in the vertebrate homeobox-genes.

**Acknowledgments**

This work is supported by Russian Foundation for Basic Researches (Grant No 96-04-49349).

**References**

1. J.M. Slack, P.W. Holland and C.F. Graham, " The zootype and the phylotypic stage" Nature **361**, 490, (1993).
2. A.V. Spirov, "The role of some conservative sequences in regulatory elements of antp-like, homeobox-containing genes of vertebrates" J. Evol. Biochem. and Physiol., [St.-Petersburg], **32**, 556 (1996)
3. B.Y. Misof and G.P. Wagner, "Evidence for four Hox clusters in the killifish Fundulus heteroclitus (teleostei)" Mol. Phylogenet. Evol., **5,** 309 (1996).
4. B.Y. Misof, M.J. Blanco and G.P. Wagner, " PCR-survey of Hox-genes of the zebrafish: new sequence information and evolutionary implications J. Exp. Zool. **274**, 193 (1996)
5. X. Li and M. Noll, "Evolution of distinct developmental functions of three Drosophila genes by acquisition of different cis-regulatory regions" Nature **367**, 83 (1994)
6. D. Reuter, R.P. Kuhnlein, G. Frommer, R. Barrio, F.C, Kafatos, H. Jackle, R. Schuh, "Regulation, function and potential origin of the Drosophila gene *spalt adjacent*, which encodes a secreted protein expressed in the early embryo" Chromosoma **104**, 445 (1996).
7. E.R. Lozovskaya, D.L. Hartl and D.A. Petrov, "Genomic regulation of transposable elements in Drosophila" Curr. Opin. Genet. Dev. **5**, 768 (1995).
8. R.J. Britten, "Mobile elements inserted in the distant past have taken on important functions" Gene, **205**, 177 (1997).
9. D. Aberdam, V. Negreanu, L. Sachs, C. Blatt, "The oncogenic potential of an activated Hox-2.4 homeobox gene in mouse fibroblasts" Mol. Cell. Biol. **11**, 554 (1991).
10. G. Vansant and W.F. Reynolds, " The consensus sequence of a major Alu subfamily contains a functional retinoic acid response element" Proc. Natl. Acad. Sci. USA **92**, 8229 (1995).
11. M.C. Moroni, M.A. Vigano and F. Mavilio, "Regulation of the human HOXD4 gene by retinoids" Mech. Dev. **44**, 139 (1993).
12. M.H. Sham, P. Hunt, S. Nonchev et al., "Analysis of the murine Hox-2.7 gene" The EMBO J. **11,** 1825 (1992).