Extreme conservation of non-repetitive non-coding regions near HoxD complex of vertebrates

Chilaka Sabarinadh, Subbaya Subramanian and Rakesh K Mishra

Addresses: Centre for Cellular and Molecular Biology, Uppal Road, Hyderabad 500 007, India.

Correspondence: Rakesh K Mishra. E-mail: mishra@ccmb.res.in

Posted: 25 March 2003

Received: 21 March 2003

© 2003 BioMed Central Ltd

This information has not been peer-reviewed. Responsibility for the findings rests solely with the author(s).

AS A SERVICE TO THE RESEARCH COMMUNITY, GENOME BIOLOGY PROVIDES A 'PREPRINT' DEPOSITORY TO WHICH ANY ORIGINAL RESEARCH CAN BE SUBMITTED AND WHICH ALL INDIVIDUALS CAN ACCESS FREE OF CHARGE. ANY ARTICLE CAN BE SUBMITTED BY AUTHORS, WHO HAVE SOLE RESPONSIBILITY FOR THE ARTICLE'S CONTENT. THE ONLY SCREENING IS TO ENSURE RELEVANCE OF THE PREPRINT TO GENOME BIOLOGY'S SCOPE AND TO AVOID ABUSIVE, LIBELLOUS OR INDECENT ARTICLES. ARTICLES IN THIS SECTION OF THE JOURNAL HAVE NOT BEEN PEER-REVIEWED. EACH PREPRINT HAS A PERMANENT URL, BY WHICH IT CAN BE CITED. RESEARCH SUBMITTED TO THE PREPRINT DEPOSITORY MAY BE SIMULTANEOUSLY OR SUBSEQUENTLY SUBMITTED TO GENOME BIOLOGY OR ANY OTHER PUBLICATION FOR PEER REVIEW; THE ONLY REQUIREMENT IS AN EXPLICIT CITATION OF, AND LINK TO, THE PREPRINT IN ANY VERSION OF THE ARTICLE THAT IS EVENTUALLY PUBLISHED. IF POSSIBLE, GENOME BIOLOGY WILL PROVIDE A RECIPROCAL LINK FROM THE PREPRINT TO THE PUBLISHED ARTICLE.
Extreme conservation of non-repetitive non-coding regions near HoxD complex of vertebrates

Chilaka Sabarinadh, Subbaya Subramanian and Rakesh K. Mishra*

Centre for Cellular and Molecular Biology
Uppal Road, Hyderabad 500 007, India

* Corresponding author

Telephone: +91 40 2719 2658
Fax: +91 40 2716 0311
Email: mishra@ccmb.res.in
Abstracts:

Homeotic gene complexes determine the anterior-posterior body axis in animals. The expression pattern and function of hox genes along this axis is colinear with the order in which they are organized in the complex. This ‘chromosomal organization and functional correspondence’ is conserved in all bilaterians investigated. Although the molecular basis of this ‘colinearity’ in not yet understood, it is possible that there are control elements within or in the proximity of these complexes that establish and maintain the expression patterns of hox genes in a coordinated fashion.

We report here an unprecedented conservation of non-coding DNA sequences adjacent to the HoxD complex of vertebrates. Stretches of hundreds of base pairs in a 7kb region, upstream of HoxD complex, show 100% conservation from fish to human. Using primers designed from these sequences of human HoxD complex, we amplified the corresponding regions from different vertebrates, including mammals, aves, reptiles, amphibians and pisces. Such a high degree of conservation, where no variation was allowed during ~500 million years of evolution, suggests critical function for these sequences in the regulation of the HoxD complex. Furthermore, these sequences provide a molecular handle to gain insight into the mechanism of regulation of this complex.
**Background:**

Eukaryotic genome contains a large excess of non-coding sequences. Conservation of these sequences among species is a strong indication of their functional significance. With the availability of genome sequences it is possible to identify such sequences taking the comparative genomics approach\(^1\)-\(^3\). Clustering of genes that are regulated in a linked manner has been noticed in several cases\(^4\),\(^5\). Among the most conserved regions of the vertebrate genome are the clusters of homeotic genes\(^6\),\(^7\). Homeotic gene complex was first identified in *Drosophila melanogaster* and was demonstrated to play major role in anterior-posterior body axis formation\(^8\). Hox genes in flies and similarly in vertebrates are expressed in a coordinated manner along the body axis. The molecular mechanism behind such coordination in regulation, however, is not yet understood. Several mechanisms have been proposed that link the organization of homeotic genes and the spatio-temporally controlled expression\(^9\)-\(^11\) of which the most attractive one implicates higher order chromatin organization in this process\(^12\). It has been shown that an upstream region spanning up to 20 kb plays an important role in the regulation of this complex\(^13\). Such studies have lead to the speculation that repressive elements in this region may initially silence the complex and then release the genes for expression in a sequential manner. Fine mapping of such sequences and their conservation in other vertebrates have not been reported. Role of higher order chromatin organization in the regulation of homeotic gene complex is relatively better known in case of bithorax complex of *Drosophila*\(^14\).

**Results and discussion:**

We compared genomic regions flanking hox complexes in order to identify conserved regions. Here we report that the upstream regions of HoxD complexes of *Homo sapiens* (human), *Mus musculus* (mouse), *Rattus norvegicus* (rat), *Papio hamadryas* (sacred baboon), *Heterodontus francisci* (horn shark), *Danio rerio* (zebra fish) and *Fugu rubripes* (Fugu) contain long stretches of extremely conserved sequences. Analysis of a 25 kb region upstream of the HoxD complex from these organisms revealed an extremely conserved region spread in three blocks located within 7 kb from the 3′ end of the *Evx-2* gene. These conserved regions, designated as Conserved Region-1, Conserved Region-2 and Conserved Region-3 (CR-1, CR-2 and CR-3) (Fig.1a) show a degree of conservation not seen before among distant species. Detailed analysis of each region spanning to
several hundred base pairs, in particular the CR-2, shows 100% conservation, Table 1, Fig.1b. These sequences are found as single copy and are vertebrate specific. We also noticed longer stretches of conservation among mammals, which gradually shortens as we go towards lower vertebrates, defining the core of each conserved region, across the vertebrate classes, Table-1. This and the fact that in case of shark, as compared to mammals, the intervening sequence lengths between CR-2 and CR-3, and CR-1 and Evx-2 is shorter by ~1300 bp and ~600 bp, respectively (Fig.1a) suggest that starting from the shorter conserved regions, additional unique sequences have progressively been acquired during the evolution of primates from lower vertebrates. We did not find such a degree of conservation in the flanking regions of other hox complexes (HoxA, B and C) of vertebrates, data not shown.

Primers designed from all three conserved regions of the human HoxD complex, amplified the corresponding regions from different species covering all five classes of vertebrates. The PCR products from different species and Southern hybridization by the human probes are shown in Fig.2. Sequencing or the PCR products confirmed these observations.

Several recent reports using comparative genomics approach have identified conserved non-coding regions among different vertebrates but none to the degree that we report here. The mechanism that may require such a high degree of conservation is not known. It is not, therefore, immediately clear what precisely is the (regulatory) role of these sequences. A part of CR-1, 2 or 3 could be the enhancer of Evx-2 gene or other regulatory elements, that could be in this region. The size and the extent of conservation, however, rules out such enhancer type regulatory sequences to be the only functional element associate with these sequences. The conserved sequences fall within the region that has been suggested to organize a repressive complex. Identification of CR-1, 2 and 3, and their ‘class specific’ extensions (Table 1) will help in the search for molecular components of any such or any other mechanism of HoxD regulation.

EST data base search revealed that part of CR-1 and CR-3 are transcribed but no EST corresponding to CR-2 or any other part of the 7.5 Kb region was found. These transcripts are expressed early in the development, Fig.1. A possible mechanism could involve RNA from this region that may be functioning by base pairing to implement temporal and spatial regulation of the homeotic genes. If that is the case, such high conservation could be expected. Role of transcription in the regulation of
bithorax complex is emerging from recent studies. Further studies will be required to determine if such a process may be common to vertebrate Hox complexes as well.

While such an extreme conservation of several hundred nucleotides over half a billion years in a region that does not code for any known proteins certainly implicates essential role for such sequences, probably in the regulation of HoxD complex, no known regulatory element requires such extreme conservation extending up to hundreds of base pairs. It is therefore, likely that these elements are a component of a novel mechanism common to all vertebrates that regulates this gene complex. We are tempted to suggest that such a strongly conserved region from fish to human linked to a gene complex that is known to determine body axis formation may be the key determinant of molecular basis of early ontogeny. Early embryos of all vertebrates show striking similarity and we suggest that these elements may be controlling the early expression pattern of HoxD which leads to similar pattern of the embryo shape. While very speculative, such possibilities can be tested experimentally. The gradient of conservation seen in this region from fish to human may signify the evolutionary history of this locus. Diversification of the vertebrate classes and the morphological features along the antero-posterior body axis that have been acquired during evolution could potentially be correlated by extensive molecular analysis of these sequences.

**Methods**

**Sequence analysis**

The genomic sequences that contained Evx-2 and any of the Hoxd genes were downloaded and annotated using gene/ORF prediction tools. Similar approach was used for other hox complexes. Homology searches of the upstream sequences of HoxD region from human (AC009336; from nucleotide 56601 to 64095) was carried out using the BLAST program of NCBI. The sequences that showed significant homology were further used to analyze the extent of homology by BLAST 2 program. The conserved regions from each sequence was obtained and subjected to multiple sequence analysis using Clustal X. In order to identify the expressed sequences corresponding to the conserved sequence, the conserved sequences along with the unique sequences were BLASTed against EST databases (human, mouse and dbEST).

The contigs that showed significant homology to the upstream sequences of human HoxD were annotated using the tBLASTx program and searching the translated amino acid sequence in the
Swissprot database. Repeat masker program was used to look for repeat content. Genebank sequences used in this study are as follows: AC116665 Papio hamadryas, AF224263 Heterodontus francisci, AC015584 Mus musculus, AC009336 Homo sapiens, CAAB01000449 Fugu rubripes and NW_042732 Rattus norvegicus.

We identified Hox complexes by searching for the respective homeotic genes and then downloading the genomic sequences. In this way we were able to study the flanking regions of HoxA, HoxB and HoxC from different vertebrate species. In order to see if there are such conserved regions associated with other complexes, we took 25kb DNA from the human HoxA, HoxB and HoxC complexes and BLASTed against ‘non redundant’ sequence and also against the genome sequence of mouse, rat, fugu and zebrafish and other eukaryotic genomes available in the public database.

**Genomic DNA isolation, PCR amplification, Sequencing and Southern hybridization:** For the isolation of genomic DNA blood samples of human, chick and cobra (Naja naja) were used while liver tissue of mouse and muscle tissue of frog (Bufo melanostictus) and zebrafish were used. Standard protocol of DNA isolation was followed which included lysis, RNase A and proteinase K digestions followed by phenol/chloroform extraction and precipitation. Concentration and quality of the genomic DNA was checked on 0.7% agarose gel and UV absorption spectrophotometry. Based on the sequence of conserved regions primers were designed to amplify the three regions CR1, CR2 and CR3. Primers used in this study to amplify conserved regions from different vertebrate species were:

CR-1 forward- GAGGCTGTTCACACTGGTGG,
CR-1 reverse– ATCATGCTCTCTGATGGACC,
CR-2 forward- GCATCGTAATCAGTTCGGTC,
CR-2 reverse- TGATACAAGCTGATACCGTC,
CR-3 forward- GCTATTCAAAATGTTATTTGAG &
CR-3 reverse- CTGTAATGAAGAAAAGATTTATG.

The 25 µl reaction was performed using 100 ng template DNA and 5 pmol each of forward and reverse primers. PCR conditions were- initial denaturation step of 94°C for 3 min was followed by 35 cycles of 94°C for 1 min, 57°C for 1 min and 72°C for 1.30 min and final extension step at 72°C for 7 min. All the PCR products were sequenced on an ABI automated DNA sequencer (Perkin Elmer)
using the ABI Big Dye terminator chemistry. For Southern hybridization, the PCR products were separated on 1% agarose gels and transferred to N\textsuperscript+ nylon membrane. Purified PCR products amplified from human DNA were labeled by random priming and used as probe. Hybridization and washings were performed as at 65\degree C. We also amplified these sequences using same primer sets from a variety of animals across the vertebrates, data not shown.

**Acknowledgements:**

We acknowledge the help from the CCMB sequencing facility, wildlife conservation group for providing DNA samples of various animal species, Mehar Sultana, A Suresh, Md Idris during this work. This work was supported by a young investigators grant (RGY0316/2001-M) from Human Frontier Science Program to RKM.

**References**

1. Pennacchio, L.A., & Rubin, EM. Genomic strategies to identify mammalian regulatory sequences. Nature Rev. Genet. 2, 100-109 (2001).
2. Kondrashov, A. S. and Shabalina, S. A. Classification of common conserved sequences in mammalian intergenic regions. Human Mol. Genetics 11, 669-674 (2002).
3. Dehal, P., Predki, P., Olsen, AS., Kobayashi, A., Folta, P., Lucas, S., Land, M., Terry, A., Ecale Zhou, C., Rash, S., Zhang, Q., Gordon, L., Kim, J., Elkin, C., Pollard, MJ., Richardson, P., Rokhsar, D., Uberbacher, E., Hawkins, T., Branscomb, E., Stubbs, L. Human Chromosome 19 and Related Regions in Mouse: Conservative and Lineage-Specific Evolution. Science 293, 104-111 (2001).
4. Boutanaev, A. M., Kalmykova, A. I., Shevelyov, Y. Y and Nuraminsky, D. I. Large clusters of co-expressed genes in the *Drosophila* genome. Nature 420, 666-669 (2002).
5. Lercher, M. J., Urrutia, A. O. and Hurst, L. D. Clustering of housekeeping genes provides a unified model of gene order in the human genome. Nat. Genetics 31, 180-183 (2002).
6. McGinnis, W. & Krumlauf, R. Homeobox genes and axial patterning. Cell 68, 283-302 (1992).
7. Krumlauf, R. *Hox* genes in vertebrate development. Cell 78, 191-201 (1994).
8. Lewis, EB. A gene complex controlling segmentation in *Drosophila*. Nature 276, 565-570 (1978).

9. Kmita, M., Kondo, T. & Duboule, D. Targeted inversion of a polar silencer within the *HoxD* complex reallocates domains of enhancer sharing. Nat. Genet. 26, 451-454, (2000).

10. Spitz, F., Gonzalez, F., Peichel, C., Vogt, TF., Duboule, D., & Zákány, J. Large scale transgenic and cluster deletion analysis of the *HoxD* complex separate an ancestral regulatory module from evolutionary innovations. Genes & Dev. 15, 2209–2214 (2001).

11. Duboule, D. Vertebrate *hox* gene regulation: clustering and/or colinearity? Curr. Opin. Genet. Dev. 8, 514-518 (1998).

12. Kmita, M., Fraudeau, N., Herault, Y. and Duboule, D. Serial deletions and duplications suggest a mechanism for the collinearity of *Hoxd* genes in limbs. Nature 420, 145-150 (2002).

13. Kondo, T. and Duboule, D. Breaking colinearity in the mouse *HoxD* complex. Cell 97, 407-417 (1999).

14. Mihaly, J., Hogga, I., Barges, S., Galloni, M., Mishra, RK., Hagstrom, K., Muller, M., Schedl, P., Siposd, L., Gausz, J., Gyurkovics, H., & Karch, F. Chromatin domain boundaries in the Bithorax complex. Cell. Mol. Life Sci. 54, 60–70 (1998).

15. Wassermann, WW., Palumbo, M., Thompson, W., 2, Fickett, JW., & Lawrence, CE. Human-mouse genome comparisons to locate regulatory sites. Nat. Genet. 26, 225-228 (2000).

16. Aparicio, S., Morrison, A., Gould, A., Gil thorpe, J., Chaudhuri, C., Rigby, PWJ., Krumlauf, R. & Brenner, S. Detecting conserved regulatory elements with the model genome of the Japanese puffer fish Fugu rubripes. Proc. Natl. Acad. Sci. USA 92, 1684-1688 (1995).

17. Dermitzakis, E. T., Reymond, A., Lyle, R., Scamuffa, N., Ucla, G., Deutsch, S., Stevenson, B. J., Flegel, V., Bucher, P., Jongeneel, C. V. and Antonarakis, S. E. Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420, 578-582 (2002).

18. Drewell, R. A., Bae, E., Burr, J. and Lewis, E. B. Transcription defines the embryonic domains of cis-regulatory activity at the *Drosophila* bithorax complex. Proc. Natl. Acad. Sci. (USA), 99, 16853-16858 (2002).

19. Rank,G., Prestel, M. and Paro, R. Transcription through intergenic chromosomal memory elements of the *Drosophila* bithorax complex correlates with an epigenetic switch. Mol. Cell. Biol. 22, 8026-8034 (2002).
20. Hogga, I. and Karch, F. Transcription through the iab-7 cis-regulatory domain of the bithorax complex interferes with the *Polycomb*-mediated silencing. Development 129, 4915-4922 (2002).

21. Bender, W. and Fitzgerald, D. P. Transcription activates repressed domain in the *Drosophila* bithorax complex. Development 129, 4923-4930 (2002).

22. Prince, V. The *hox* paradox: more complex(es) than imagined. Dev. Biol. 249, 1-15 (2002).

23. Manzanares, M., Wada, H., Itasaki, N., Trainor PA, Krumlauf, R. & Holland, P. W. Conservation and elaboration of *Hox* gene regulation during evolution of the vertebrates. Nature 408, 854-857 (2000).
Table 1. Conservation of sequences in the regions CR-1, 2 and 3 in different vertebrate species.

| Comparison              | CR-1 Core | CR-1 Extended | CR-2 Core | CR-2 Extended | CR-3 Core | CR-3 Extended |
|-------------------------|-----------|---------------|-----------|---------------|-----------|---------------|
| human                   | 317 [99]  | 3119 [95]     | 800 [99.4]| 4400 [95]     | 256 [99.6]|               |
| included in baboon      |           |               |           |               |           |               |
| human/rat               | 317 [98.4]| 837 [91]      | 800 [97.8]| 1127 [95]     | 256 [99.2]| 748 [94]      |
| human/mouse             | 317 [98.4]| 846 [92]      | 800 [97.8]| 568 [98]      | 256 [99.6]| 652 [97]      |
| mouse/rat               | 317 [98.7]| 2180 [95]     | 800 [98]  | 1879 [92]     | 256 [99.6]| 855 [81]      |
|                         |           |               |           | 964 [94]      |           | 483 [90]      |
| mouse/rat               | 317 [98.7]| 2180 [95]     | 800 [98]  | 1879 [92]     | 256 [99.6]| 855 [81]      |
|                         |           |               |           | 964 [94]      |           | 483 [90]      |
| vertebrate              | 43, 30 [100]| 144, 94, 59, | 91, 62, 35,|               |           |               |
|                         | 197 [93]  | 53, 46 [100]  | 62 [100]  |               |           |               |
|                         |           | 463 [96]      |           |               |           |               |

Vertebrate conservation is based on human, mouse and shark comparison. Although we find genomic sequences of other vertebrates (fugu and zebrafish) that show conservation in this region, those sequences are in draft form and, therefore, have been excluded from this comparison. Length of core and extended regions of conservation are given in base pairs. Number within the bracket indicates percent homology.
**Figure 1a.** Schematic representation of sequence conservation in the *HoxD* upstream region. Sequences that are conserved across vertebrates are shown as blocks. The conservation extends beyond these blocks with in primates and rodents. ESTs found in the database corresponding to this region are also shown. ESTs mapping to CR-3 are BB838602 from mouse 8 cell embryo and BU129154 from chicken 36 stage limb; and those mapping to CR-1 are AA620964 from human testis; BB332383, BB335110, BB334358, BB333569 from 6 and 10 days mouse neonate medulla oblongata and BU255316 from chicken 36 stage limb.
Figure 1b. Comparison of conserved regions from human, mouse and shark. Conserved bases of mouse and shark are shown as '.' and '-' indicates indels. Underlined sequences of human indicate primers that were used for amplification of the corresponding sequence from different vertebrates.
**Figure 2.** PCR amplification and the southern hybridization of different vertebrate genomic DNA samples using primers designed based on human sequence. Lanes 1-6 represent human, mouse, chick, cobra, frog and fish samples, respectively. Probe used in hybridization is made of corresponding human amplicons.