Phylogeny of the Insect Homeobox Gene (*Hox*) Cluster

Sangeeta Dhawan and K. P. Gopinathan*

Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore 560012, India.

The homeobox (*Hox*) genes form an evolutionarily conserved family encoding transcription factors that play major roles in segmental identity and organ specification across species. The canonical grouping of *Hox* genes present in the HOM-C cluster of *Drosophila* or related clusters in other organisms includes eight "typical" genes, which are localized in the order *labial* (*lab*), *proboscipedia* (*pb*), *Deformed* (*Dfd*), *Sex combs reduced* (*Scr*), *Antennapedia* (*Antp*), *Ultrabithorax* (*Ubx*), *abdominalA* (*abdA*), and *AbdominalB* (*AbdB*). The members of *Hox* cluster are expressed in a distinct anterior to posterior order in the embryo. Analysis of the relatedness of different members of the *Hox* gene cluster to each other in four evolutionarily diverse insect taxa revealed that the loci *pb/Dfd* and *AbdB*, which are farthest apart in linkage, had a high degree of evolutionary relatedness, indicating that *pb/Dfd* type anterior genes and *AbdB* are closest to the ancestral anterior and posterior *Hox* genes, respectively. The greater relatedness of other posterior genes *Ubx* and *abdA* to the more anterior genes such as *Antp* and *Scr* suggested that they arose by gene duplications in the more anterior members rather than the posterior *AbdB*.

Key words: evolution, homeodomain, *Hox* genes, insects, phylogeny

Introduction

Homeotic (*Hox*) genes encode highly conserved, homeodomain-containing transcription factors and are important developmental regulators that act together to determine segment identity (*1*, *2*). Each *Hox* gene controls the expression of a variety of target genes (*3*, *4*). A shift in the expression pattern of a *Hox* gene may lead to altered morphology and thus providing a mechanism of relatively rapid macroevolutionary change (*5*).

The *Hox* genes are typically found together in a single complex on the chromosome and promote the identity of segments along the anterior-posterior axis of the embryo in the same order in which they lie on the chromosome (*2*, *6*), a phenomenon termed "Colinearity rule" (*1*). The members of *Hox* cluster are expressed in a distinct anterior to posterior order in the embryo and on this basis are classified as anterior-, middle-, or posterior expressing genes. The canonical grouping of *Hox* genes pertains to the genes present in the HOM-C cluster of *Drosophila* and related clusters in other organisms. Other homeobox genes that play important roles in segment identity have also been considered honorary *Hox* genes and may have been linked with the *Hox* cluster in the past (*e.g.*, *eve*; ref. *7*). The *Hox* cluster is formed by ten genes: the eight "typical" *Hox* genes [*labial* (*lab*), *proboscipedia* (*pb*), *Deformed* (*Dfd*), *Sex combs reduced* (*Scr*), *Antennapedia* (*Antp*), *Ultrabithorax* (*Ubx*), *abdominalA* (*abdA*), and *AbdominalB* (*AbdB*)] and the two "atypical" genes: *Hox 3* (*zen*, *z2* and *bicoid*) and *fushi tarazu* (*ftz*). The atypical genes do not play *Hox*-like roles in *Drosophila* but appear to function as *Hox* genes in more basal arthropods (*8–12*).

One of the central questions in *Hox* gene evolution has been to figure out how the various members of the *Hox* cluster arose and diverged from the ancestral *Hox* genes. The structure of the *Hox* gene cluster and the functions of some *Hox* genes have undergone subtle changes in different insect taxa and provide an interesting system to analyze the *Hox* gene evolution. In this study, we attempt to present a model for the evolution of the *Hox* gene cluster in the insects based on the phylogenetic relationship between different members of the insect *Hox* family. The phylogenetic relatedness of various members of the *Hox* cluster was examined in four divergent insect models. In these analyses, the most posterior gene *AbdB* was found to be phylogenetically more related to the very anterior members of the clusters *pb* and *Dfd*, rather than to

*** Corresponding author.**
E-mail: kpg@mcbl.iisc.ernet.in

the other posterior genes such as *Ubx* and *abdA*. This result suggested that *pb*/*Dfd* type anterior genes and *AbdB* are closest to the ancestral anterior and posterior *Hox* genes, respectively. The greater relatedness of other posterior genes such as *Ubx* and *abdA* to more anterior genes like *Antp* and *Scr* suggested that these posterior genes arose by gene duplications in the more anterior members, rather than the posterior *AbdB*.

Results and Discussion

The *Hox* genes are linked in the order *lab*, *pb*, *Dfd*, *Scr*, *Antp*, *Ubx*, *abdA*, and *AbdB*, in the prototype model *Drosophila*. The members of the *Hox* cluster share a great degree of sequence homology. To analyse the *Hox* evolution in insects, molecular phylogenies were constructed for comparing the different members of *Hox* cluster from four representative insect taxa, *Bombyx mori* (Lepidopteran), *Drosophila melanogaster* (Dipetran), the beetle *Tribolium castaneum* (Coleoptrean), and the fire-brat *Thermobia domestica* (Thysanuran), with *Thermobia* and *Drosophila* being the most basal and derived species respectively, amongst the ones analyzed here.

The *Hox* genes encode proteins containing the homeodomain, a highly conserved functional domain that binds to the DNA, with *AbdB* having the most divergent homeodomain sequence amongst all. Phylogenetic analysis was carried out using the homeodomain encoding nucleotide sequences for the *Hox* cluster members from the four representative insect taxa by Neighbour-joining analysis (Table 1). Use of nucleotide sequences was preferred over the protein sequences as this allows one to monitor and compare even conserved sequence variations in the sequences, leading to a more sound analysis of sequence divergence in a highly conserved domain. In each case, the genes *Ubx* and *abdA* (*Utx* and *A* in *Tribolium*) clustered together, indicating a common origin. Also *Scr* and *Antp* (*Cx* and *ptl* in *Tribolium*) were closely related in all the cases examined. *Dfd* and *Scr* also showed a high degree of relatedness to each other. These results suggested a possible common origin for *Dfd*, *Scr* and *Antp*. Although *lab* is the anterior-most gene in these clusters, it was closer to *Scr* and *Antp* than the closest neighbour on the cluster *pb*. Surprisingly, the second most anterior gene in the cluster *pb* always grouped with the most posterior gene *AbdB* (in case of *Thermobia*, *Tribolium* and *Drosophila*) (Figure 1A-D). Since *pb* homologue has not been identified so far in *Bombyx*, the relatedness of *pb* to *AbdB* cannot be commented upon in this instance. However, in *Bombyx*, the anterior class gene *Dfd* was found to be closer to the most posterior *AbdB* than to the anterior class genes (Figure 1C). In *Thermobia*, *Dfd* was also found to group with *AbdB* (Figure 1A). These observations were further confirmed by constructing a phylogeny of nucleotide sequences encoding homeodomains from all the known *Hox* genes from the same four insect taxa (Figure 1E). To further substantiate these findings, a protein-sequence-based phylogenetic analysis was performed using the 60-amino-acid-long homeodomain sequence from *Bombyx*, *Drosophila* and *Tribolium* (*Thermobia* was not used in this analysis because of lack of complete sequence information within the homedomain for some *Hox* genes in this organism) (Figure 2). Phylogenetic analysis was performed using Minimum evolution test with bootstrap analysis. Even in the protein based analysis, the *AbdB* sequences were found to cluster with the *pb* and *lab* sequences, indicating the evolutionary relatedness of *AbdB* to the extreme anterior *Hox* genes.

Table 1 GenBank Accession Numbers of the Nucleotide Sequences Used for Homeodomain Phylogenetic Analysis

Gene	*Bombyx mori*	*Drosophila melanogaster*	*Tribolium castaneum*	*Thermobia domestica*
lab	NA	X12834	AF231104	AF104008
pb	NA	X63728	AF187068 (*mxp*)	AF104009
Dfd	D83534	X05136	U81038	AF104005
Scr	D83533	X14475	AF227628 (*Cx*)	AF104010
Antp	D16684	X03791	AF228509 (*ptl*)	AF104003
Ubx	X62618	X76210	AF146650 (*Utx*)	NA
abdA	X62620	X54453	X72339 (*A*)	AF104001
AbdB	X62619	X51663	AF227923	AF104002

The protein phylogenies were constructed based on the amino acid translate sequence data obtained from the sequences listed here.

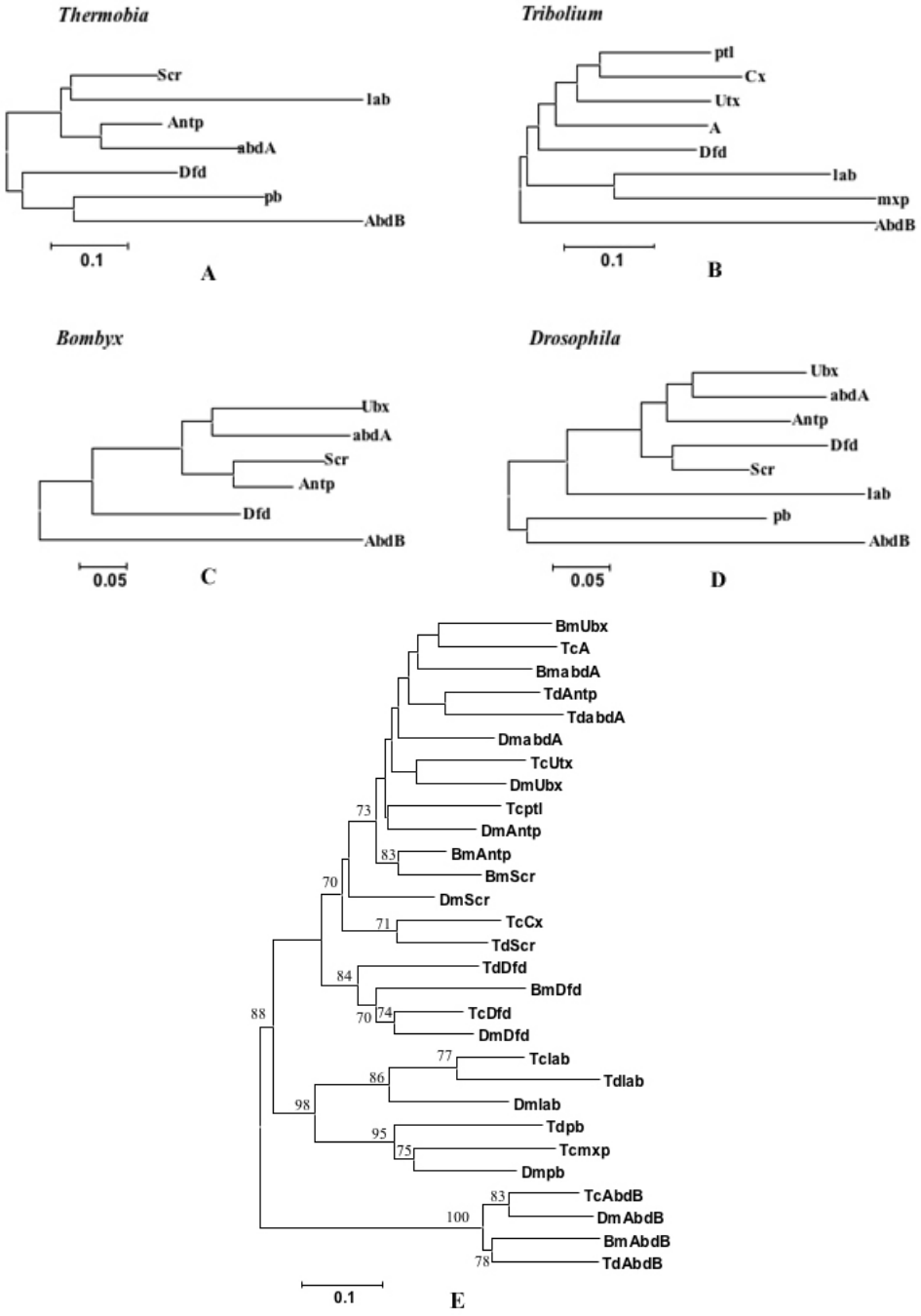

Fig. 1 Evolutionary relatedness of various members of *Hox* cluster. **A–D.** Phylogenetic analysis of *Hox* cluster from four representative insect taxa. Panels depict *Hox* gene phylogenies for *Thermobia* (A), *Tribolium* (B), *Bombyx* (C), and *Drosophila* (D). The phylogenetic trees were constructed by Neighbour-joining method as implemented in MEGA2 software (*17*), with bootstrap analysis. All the nodes represented a bootstrap support value above 50. The sequences used and their GenBank accession numbers are listed in Table 1. The *Tribolium Hox* homologues have been depicted with their classical nomenclature. *mxp* (*maxillapedia*): *pb* homologue; *Cx* (*Cephalothorax*): *Scr* homologue; *ptl* (*prothoraxless*): *Antp* homologue; *Utx* (*Ultrathorax*): *Ultrabithorax* homologue; and *A* (*Abdominal*): *abdA* homologue. The scale bar indicates the distance between different sequences. **E.** A combined tree of all the available members of *Hox* clusters from the four insect taxa was constructed using the Neighbour-joining method as implemented in the MEGA3 software (*19*) with bootstrap analysis. The bootstrap values of various nodes are marked. The nodes with no value displayed had low support values (below 70). The scale bar indicates the distance between various sequences. Note the grouping of *AbdB* with *pb* and *lab* group of anterior genes.

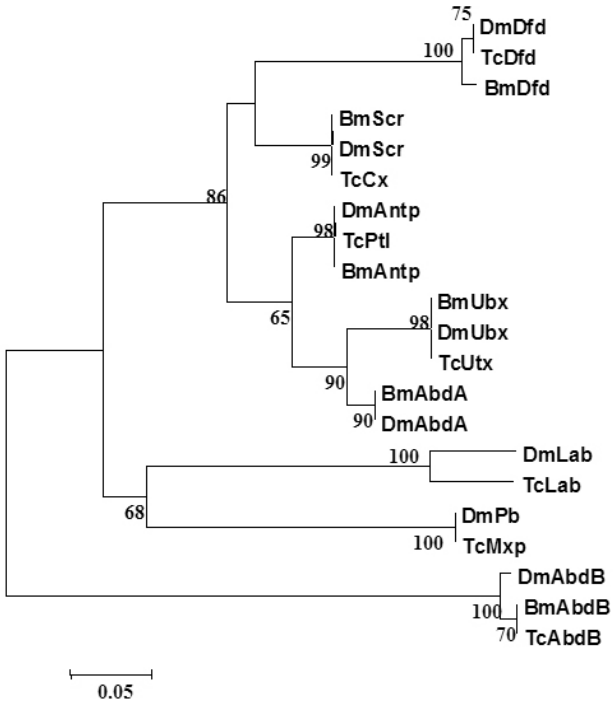

Fig. 2 Protein-sequence-based analysis of the evolutionary relatedness of various homedomains. Panels depict *Hox* gene phylogenies for *Tribolium*, *Bombyx* and *Drosophila*. A combined tree of all the available members of classical Hox proteins from these three insect taxa was constructed by Neighbour-joining method as implemented in MEGA3 software (*19*), with bootstrap analysis. The sequences used were derived by translating the data from sequences listed in Table 1. The *Tribolium Hox* homologues have been depicted with their classical nomenclature. *mxp* (*maxillapedia*): *pb* homologue; *Cx* (*Cephalothorax*): *Scr* homologue; *ptl* (*prothoraxless*): *Antp* homologue; *Utx* (*Ultrathorax*): *Ultrabithorax* homologue; and *A* (*Abdominal*): *abdA* homologue. The scale bar indicates the distance between different sequences. The bootstrap values of various nodes are marked. The nodes with no value displayed had low support values (below 70). All the nodes represented a bootstrap support value above 50. The scale bar indicates the distance between various sequences. Note the grouping of *AbdB* with *pb* and *lab* group of anterior genes.

Our analyses suggested a common ancestor for *pb* and *AbdB*, as well as *Dfd*. Since the other posterior genes *Ubx* and *abdA* were closer to *Antp* and the other central and anterior genes than to *AbdB*, it may be concluded that *AbdB* is the only ancestral posterior class gene and the rest were derived by duplications of the anterior and central genes.

Based on the known distribution of *Hox* clusters, the origin of the *Hox* cluster is thought to have predated the radiation of triploblastic metazoans, the bilateria. In most bilaterians surveyed, several distinct *Hox* gene subsets have been found. They are designated as the "head", "trunk", and "tail" genes or the 5′, central, and 3′ genes, depending on their patterns of expression across the embryonic axis or their distributions in the *Hox* cluster. Most likely, the ancestral bilaterian possessed a *Hox* cluster consisting of two anterior members, a posterior member and possibly a central member. The subsequent variety in the *Hox* cluster then arose presumably by gene duplication events (*13*). The pertinent questions then are, how the various members of the cluster are related to each other, and what is ancestral and what is derived. Molecular phylogenies provide a valuable tool in answering these questions. A strong phylogenetic relationship was noticed between the most posterior gene *AbdB* and two anterior genes *pb* and *Dfd*, rather than to the immediate neighbouring genes of *AbdB*, namely *abdA* and *Ubx*. These results suggested two possibilities: (a) the *pb*/*Dfd* type anterior genes and *AbdB* are more like the ancestral anterior and posterior *Hox* genes, as evidenced by their relatedness; and (b) the other posterior genes like *abdA* and *Ubx* arose by gene duplications in the more anterior members, as seen from the relatedness of these two genes to *Scr* and *Antp*, or they possibly arose from *AbdB* very early and underwent large divergences in the sequence.

Materials and Methods

Sequence homology searches

Sequence homology searches were performed using BLAST (basic local alignment search tool) server at NCBI (*14*), or WU BLAST and FASTA servers at EBI, UK. Published sequence data was accessible from the GenBank database at NCBI. All the novel sequences reported in this study were also deposited to GenBank. Alignments of DNA and protein sequences were generated using CLUSTAL W 1.82 (*15*) and MultiAlin (*16*) software. For CLUSTAL W, alignments were obtained using the BLOSUM matrix, a gap-opening penalty of 10 and a gap-extension penalty of 0.2.

Phylogenetic analysis of nucleic acids and protein sequences

The phylogenetic analysis of various sequences was performed using MEGA2 program (*17*) or PAUP Version 4.0 (*18*), with various phylogenetic analysis protocols such as "Minimum evolution", "Neighbour-joining", or "Quartet-puzzling".

Acknowledgements

We thank the Department of Biotechnology, Government of India, for financial support. SD was a recipient of a Senior Research Fellowship from CSIR (Council of Scientific and Industrial Research), New Delhi, and KPG is an INSA (Indian National Science Academy) Senior Scientist.

References

1. Lewis, E.B. 1978. A gene complex controlling segmentation in *Drosophila*. *Nature* 276: 565-570.
2. Kaufman, T.C., *et al.* 1990. Molecular and genetic organization of the antennapedia gene complex of *Drosophila melanogaster*. *Adv. Genet.* 27: 309-362.
3. Mastick, G.S., *et al.* 1995. Identification of target genes regulated by homeotic proteins in *Drosophila melanogaster* through genetic selection of Ultrabithorax protein-binding sites in yeast. *Genetics* 139: 349-363.
4. Botas, J. and Auwers, L. 1996. Chromosomal binding sites of Ultrabithorax homeotic proteins. *Mech. Dev.* 56: 129-138.
5. Hughes, C.L. and Kaufman, T.C. 2002. *Hox* genes and the evolution of the arthropod body plan. *Evol. Dev.* 4: 459-499.
6. Sanchez-Herrero, E. *et al.* 1985. Genetic organization of *Drosophila* bithorax complex. *Nature* 313: 108-113.
7. Valentine, J.W., *et al.* 1996. Developmental evolution of metazoan body plans: the fossil evidence. *Dev. Biol.* 173: 373-381.
8. Damen, W.G. and Tautz, D. 1998. A *Hox* class 3 orthologue from the spider *Cupiennius salei* is expressed in a *Hox*-gene-like fashion. *Dev. Genes Evol.* 208: 586-590.
9. Telford, M.J. and Thomas, R.H. 1998. Of mites and *zen*: expression studies in a chelicerate arthropod confirm *zen* is a divergent *Hox* gene. *Dev. Genes. Evol.* 208: 591-594.
10. Gibson, G. 2000. Evolution: *Hox* genes and the cellared wine principle. *Curr. Biol.* 10: R452-455.
11. Alonso, C.R., *et al.* 2001. Evolution of Ftz protein function in insects. *Curr. Biol.* 11: 1473-1478.
12. Lohr, U., *et al.* 2001. *Drosophila fushi tarazu*: a gene on the border of homeotic function. *Curr. Biol.* 11: 1403-1412.
13. Finnerty, J.R. and Martindale, M.Q. 1998. The evolution of the *Hox* cluster: insights from outgroups. *Curr. Opin. Genet. Dev.* 8: 681-687.
14. Altschul, S.F., *et al.* 1990. Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.
15. Thompson, J.D., *et al.* 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighing, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673-4680.
16. Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* 16: 10881-10890.
17. Kumar, S., *et al.* 2001. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17: 1244-1245.
18. Swofford, D.L. 2002. PAUP*: phylogenetic analysis using parsimony (and other methods). Version 4 beta 10. CD-ROM. Sinauer Associates, Inc., Publishers, Sunderland, Massachusetts, USA.
19. Kumar, S., *et al.* 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* 5: 150-163.