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...
The protein phylogenies were constructed based on the amino acid translate sequence data obtained from the sequences listed leading to a more sound analysis of sequence divergence. 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 (Dipteran), 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, 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

<table>
<thead>
<tr>
<th>Gene</th>
<th>Bombyx mori</th>
<th>Drosophila melanogaster</th>
<th>Tribolium castaneum</th>
<th>Thermobia domestica</th>
</tr>
</thead>
<tbody>
<tr>
<td>lab</td>
<td>NA</td>
<td>X12834</td>
<td>AF231104</td>
<td>AF104008</td>
</tr>
<tr>
<td>pb</td>
<td>NA</td>
<td>X63728</td>
<td>AF187068 (mux)</td>
<td>AF104009</td>
</tr>
<tr>
<td>Dfd</td>
<td>D83534</td>
<td>X05136</td>
<td>U81038</td>
<td>AF104005</td>
</tr>
<tr>
<td>Scr</td>
<td>D83533</td>
<td>X14475</td>
<td>AF227628 (Cz)</td>
<td>AF104010</td>
</tr>
<tr>
<td>Antp</td>
<td>D16684</td>
<td>X03791</td>
<td>AF228509 (ptl)</td>
<td>AF104003</td>
</tr>
<tr>
<td>Ubx</td>
<td>X62618</td>
<td>X76210</td>
<td>AF146650 (Utx)</td>
<td>NA</td>
</tr>
<tr>
<td>abdA</td>
<td>X62620</td>
<td>X54453</td>
<td>X72339 (A)</td>
<td>AF104001</td>
</tr>
<tr>
<td>AbdB</td>
<td>X62619</td>
<td>X51663</td>
<td>AF227923</td>
<td>AF104002</td>
</tr>
</tbody>
</table>

The protein phylogenies were constructed based on the amino acid translate sequence data obtained from the sequences listed here.
Evolution of Hox Genes in Insects

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. maxp (maxillapedia); pb homologue; Cx (Cephalothorax); Scr homologue; pti (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.
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