Hox genes and evolution of body plan

Prof. LS Shashidhara
Indian Institute of Science Education and Research (IISER), Pune
ls.shashidhara@iiserpune.ac.in
2009 marks 150 years since Darwin and Wallace proposed theory of natural selection and also marks bicentenary of Darwin’s birth.

According to natural selection there is continuous interaction between changing genetic architecture of living organisms with changing habitat/environment and this leads to formation of myriad of different kinds of species.
While enthusiastic Darwinists used popular phrases “struggle for existence” and “survival of the fittest” to dramatize his theory of natural selection, it means survival of those which have genetic variations that are appropriate for a given environment.
According to the widely accepted theory of natural selection, the whole process is blind. Genetic variations occur randomly and their selection by nature is purely based on their adaptability in given time and space.
Evolution means *change*

- Evolution does not mean progress or improvement.

- It is a process of adaptation to survive in a constantly changing environmental condition.
During development a multicellular organism develops from a unicellular embryo.
Morphological events are preceded by molecular events.
Hox genes regulate segment specific developmental pathways
Hox genes specify body plan

Wild Type

Antp Mutant
Hox genes specify body plan
Comparison of Hox genes in fly and mouse embryos

Organization of Hox genes, their sequences and function – all are conserved from flies to mice to human.
Normal mouse
13T + 6L + 4S
Hox10 expression
Lumbar to posterior
Hox11 expression
Sacral to posterior

Hox 10 & 11 mutant
23 Thoracic vertebrae
No lumbar
No sacral
Several models linking Hox evolution to changes in adult body plan

- Changes in the number of Hox gene (duplication and divergence)
- Changes in domain of Hox gene expression
- Changes in Hox gene that gives the protein new properties
- Changes in Hox-protein responsive elements of downstream genes
Figure 1 Genome-wide comparison of transcriptional activator families in eukaryotes. The relative sizes of transcriptional activator families among *Homo sapiens*, *D. melanogaster*, *C. elegans* and *S. cerevisiae* are indicated, derived from an analysis of eukaryotic proteomes using the INTERPRO database, which incorporates Pfam, PRINTS and Prosite. The transcription factors families shown are the largest of their category out of the 1,502 human protein families listed by the IPI.
Duplication of conserved Hox gene cluster during evolution

5' end of cluster expressed late in posterior body parts

3' end of cluster expressed early in anterior body parts

Copyright 2003 John Wiley and Sons, Inc.
Fore Limb Vs Hind Limb

Tbx5

Hoxc4 Hoxc5

Hoxc6 Hoxc8

Hoxc9 Hoxc10 Hoxc11

Ptx1

Tbx4
Several models linking Hox evolution to changes in adult body plan

- Changes in the number of Hox gene (duplication and divergence)

- Changes in domain of Hox gene expression

- Changes in Hox gene that gives the protein new properties

- Changes in Hox-protein responsive elements of downstream genes
Arthropod Phylogeny

- **CHELICERATA** (scorpions, spiders, horseshoe crabs)
- **MYRIAPODA** (centipedes, millipedes)
- **CRUSTACEA** (lobsters, shrimp, etc.)
  - Odonata (dragonflies, etc.)
  - Orthoptera (grasshoppers, etc.)
  - Coleoptera (beetles)
  - Lepidoptera (butterflies, moths)
  - Diptera (flies, mosquitoes)

**INSECTA**
Fore Limb Vs Hind Limb

Tbx5

Hoxc4 Hoxc5

Hoxc6 Hoxc8

Hoxc9 Hoxc10 Hoxc11

Ptx1

Tbx4
Several models linking Hox evolution to changes in adult body plan

- Changes in the number of Hox gene (duplication and divergence)
- Changes in domain of Hox gene expression
- Changes in Hox gene that gives the protein new properties
- Changes in Hox-protein responsive elements of downstream genes
Phylum Arthropoda

Velvet worm
*Class Onycophora*

Common brine shrimp (*Artemia*)
*Class Crustacea*

Butterfly
*Class Insecta*
*Order Lepidoptera*

*Drosophila*
*Class Insecta*
*Order Diptera*
Poly-alanine rich and glutamine/alanine rich sequences found in many repression domains.

Mediate repression by interacting with basal transcriptional machinery.
DUbx

N terminal OUbx

OUbx /QA
Evolution of insect Ubx protein by loss of CK11 and GSK phosphorylation sites and expansion of the QA domain, thus contributing to hexapod body plan.
Change in Ubx protein
Several models linking Hox evolution to changes in adult body plan

- Changes in the number of Hox gene (duplication and divergence)
- Changes in domain of Hox gene expression
- Changes in Hox gene that gives the protein new properties
- Changes in Hox-protein responsive elements of downstream genes
Downstream of Homeotic genes...

No $Ubx$ in T3

Wildtype

$Ubx$ in both T2 and T3

Organ identity: wing vs haltere
homeotic mutant

dragonfly
over-expression of Ubx from butterflies and even from a non-winged arthropod such as Onychophora is sufficient to induce wing-to-haltere transformations in *Drosophila*

The difference must lie in the response of the downstream targets of Ubx.
In butterflies...

_Hind sight_ mutants exhibit similarity between fore and hind wings

_Ubx_ is required for hind wing identity in butterflies
During dipteran evolution, certain wing-patterning genes must have come under the regulation of Ubx.
Downstream of Homeotic genes...

Organ identity: wing vs haltere

No *Ubx* in T3

Wildtype

*Ubx* in both T2 and T3
Summary,

Ubx specifies haltere fate by down-regulating key signal transduction pathways, such as Wnt, Dpp and EGFR.

*Mechanism of Development (GEP)* 5, 113-121 (2004)
The difference must lie in the response of the downstream targets of Ubx
Several models linking Hox evolution to changes in adult body plan

- Changes in the number of Hox gene (duplication and divergence)
- Changes in domain of Hox gene expression
- Changes in Hox gene that gives the protein new properties
- Changes in Hox-protein responsive elements of downstream genes
During dipteran evolution, certain wing-patterning genes must have come under the regulation of *Ubx*.
Identification of Dipteron-specific targets of Ubx.

ChIP on different insect groups such as Apis, Butterflies, silkworm, Tribolium, mosquito and (at least two species of) Drosophila.

Bioinformatics analyses
All these genomes have been sequenced, which enables global ChIP experiments for identifying targets of Ubx.
ChIP-chip: A High throughput method to identify binding sites for any Transcription Factor

Agilent Technologies, Inc.
Santa Clara, CA 95051
• Polyclonal Antibodies generated against N-terminal region of Drosophila Ubx

Unique N-terminal region

Conserved Homeo-domain
Post ChIP-chip

1. Validation by RNA in situ, q-PCR (independent of Ubx polyclonal antibodies)

2. Functional characterization of some interesting candidate genes

3. Data Mining
Overview of Analysis Strategy to find Motifs


Start with 519 sequences → Remove coding regions → Remove sequences less than 8 characters

Filter results through CLOVER

Run MEME over sets of 35 sequences → Break into random sets of 35 sequences

Use STAMP to cluster results

Compare with TF databases like JASPAR, Biobase to identify known/novel motifs

Wet lab validation
Hox Paradox

UbxA core binding sequence TAAT is a common binding site for many other Hox genes in-vitro (Lohmann I. et al., 2008).

How Hox proteins select specific targets in vivo?

Lohmann I. et al., 2008
Motifs found from probes with p<0.01 (255 Probes)

<table>
<thead>
<tr>
<th>Motif</th>
<th>#Occur (%)</th>
<th>Consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Motif" /></td>
<td>78(30.5)</td>
<td>ACAC[AG]C[AG]CACAC</td>
</tr>
<tr>
<td><img src="image2" alt="Motif" /></td>
<td>64(25.1)</td>
<td>CTCTC[CT]CTCTC</td>
</tr>
<tr>
<td><img src="image3" alt="Motif" /></td>
<td>82(32.2)</td>
<td>[GA]CA[AG]CAACAAACA</td>
</tr>
<tr>
<td><img src="image4" alt="Motif" /></td>
<td>63(24.7)</td>
<td>GCAG[AC][GC]GCAGC</td>
</tr>
<tr>
<td><img src="image5" alt="Motif" /></td>
<td>66(25.9)</td>
<td>[CT]CCC[TAC][CTA][CT] [CA]CC[CA]C</td>
</tr>
<tr>
<td><img src="image6" alt="Motif" /></td>
<td>51(20.0)</td>
<td>G[ACG]GA[GA][AC]GAG[CA]G[CA]</td>
</tr>
</tbody>
</table>
Data from ChIP-chip

519 probes after cutoff for enrichment and p-value

Add 500 bp on either side of the probes

Extract aligning sequences from related species
Comparison of TFs in two-winged vs those in 4-winged Apis

BRCZ3_01, HB_01, BYN_Q6, SD_Q6, DL_02, BCD_01, CF2II_02, BRCZ2_01, OVO_Q6, UBX_01, SN_02, EN_Q6, ABDA_Q6, BRCZ1_01, CEBP_Q6, ANTF_Q6_01, MAD_Q6, ABDB_Q6, CAD_Q6, DEAF_01, CF2II_01, DREF_Q3, BRK_Q6, CROC_01, ZEN_Q6, FTZ_01, SGF3_Q6, PRD_Q6, TCF_Q6, ADF1_Q6

MTTFA_01, ABDB_Q6, BRCZ3_01, HB_01, CAD_Q6, DEAF_01, CF1A_Q6, ZEN_Q6, SD_Q6, DL_01, CROC_01, DL_02, CF2II_01, CF2II_02, DRI_01, FTZ_01, GRH_01, SGF3_Q6, UBX_01, SN_02, PRD_Q6, BCD_01, TCF_Q6, ABDA_Q6, ANTP_Q6_01
Detecting regulatory TFs using homology between different species

• Case Study:
  – *pipsqueak*
TRANSFAC analysis

• Take region 2kb upstream of pipsqueak in \textit{D melanogaster} and regions aligning to it, from \textit{D pseudoobscura}, \textit{D virilis}, \textit{A mellifera}, and \textit{T castaneum}

• Locate TFBSs in each sequence using the TRANSFAC Pro database (66 insect TF insect)
TFBSs (from TRANSFAC) between insects in the promoter region (1.5kb upstream) of *pipsqueak*

<table>
<thead>
<tr>
<th></th>
<th>Dmel</th>
<th>Dpse</th>
<th>Dvir</th>
<th>Amel</th>
<th>Tcas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dmel</td>
<td>15(34)</td>
<td>8</td>
<td>11</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Dpse</td>
<td>7</td>
<td>13(28)</td>
<td>8</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Dvir</td>
<td>4</td>
<td>5</td>
<td>16(30)</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Amel</td>
<td>9</td>
<td>6</td>
<td>9</td>
<td>16(34)</td>
<td>6</td>
</tr>
<tr>
<td>Tcas</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>8(17)</td>
</tr>
</tbody>
</table>

**Common** **Difference (top-left)** **TFs (total sites)**
Next slide compares all TFs in two-winged vs those in 4-winged

- **Apis mel**
  - BRCZ4_01, ABDB_Q6, CAD_Q6, HB_01, BRCZ3_01, DEAF_01, TWI_Q6, ABDB_01, CROC_01, CF2II_01, CF2II_02, DRI_01, FTZ_01, SGF3_Q6, PRD_Q6, CEBP_Q6, ABDA_Q6, TCF_Q6

- **T cast**
  - MAD_Q6, ADF1_Q6, ABDB_Q6, BRCZ3_01, HB_01, CAD_Q6, DEAF_01, SD_Q6, ABDB_01, BRK_Q6, DL_02, EVE_Q6, ZEN_Q6, DRI_01, FTZ_01, SGF3_Q6, UBX_01, SN_02, PRD_Q6, ABDA_Q6, BRCZ1_01, GAGAFACTOR_Q6, ANTP_Q6_01
What is the selection pressure for the evolution of two-winged insects?

<table>
<thead>
<tr>
<th></th>
<th>Wingbeats per second</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Odonata</strong></td>
<td></td>
</tr>
<tr>
<td>Libellula</td>
<td>20</td>
</tr>
<tr>
<td>Aeshna</td>
<td>22, 28</td>
</tr>
<tr>
<td><strong>Coleoptera</strong></td>
<td></td>
</tr>
<tr>
<td>Coccinella</td>
<td>75-91</td>
</tr>
<tr>
<td>Melolontha</td>
<td>46</td>
</tr>
<tr>
<td><strong>Lepidoptera</strong></td>
<td></td>
</tr>
<tr>
<td>Pieris</td>
<td>9, 12</td>
</tr>
<tr>
<td>Saturnia</td>
<td>8</td>
</tr>
<tr>
<td>Macroglossa</td>
<td>72, 85</td>
</tr>
<tr>
<td>Papilio</td>
<td>5-9</td>
</tr>
<tr>
<td><strong>Diptera</strong></td>
<td></td>
</tr>
<tr>
<td>Aedes (male)</td>
<td>587</td>
</tr>
<tr>
<td>Culex</td>
<td>278-307</td>
</tr>
<tr>
<td>Musca</td>
<td>190, 180-197</td>
</tr>
<tr>
<td>Tabanus</td>
<td>96</td>
</tr>
<tr>
<td>Forcipomyia</td>
<td>988-1047</td>
</tr>
<tr>
<td><strong>Hymenoptera</strong></td>
<td></td>
</tr>
<tr>
<td>Apis</td>
<td>190, 250</td>
</tr>
<tr>
<td>Bombus</td>
<td>130, 240</td>
</tr>
<tr>
<td>Vespa</td>
<td>110</td>
</tr>
</tbody>
</table>
Drosophila beats its wings about 200 times per second.

CREDIT: Michael Dickinson
Vorticity in dragonfly sequence

Vorticity in a dipteran sequence