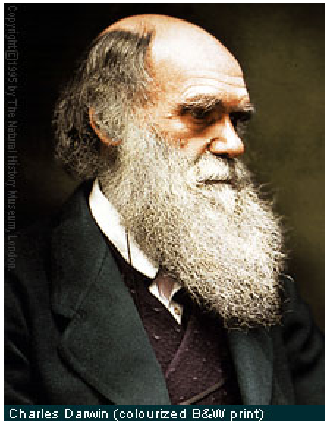


Hox genes and evolution of body plan

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.

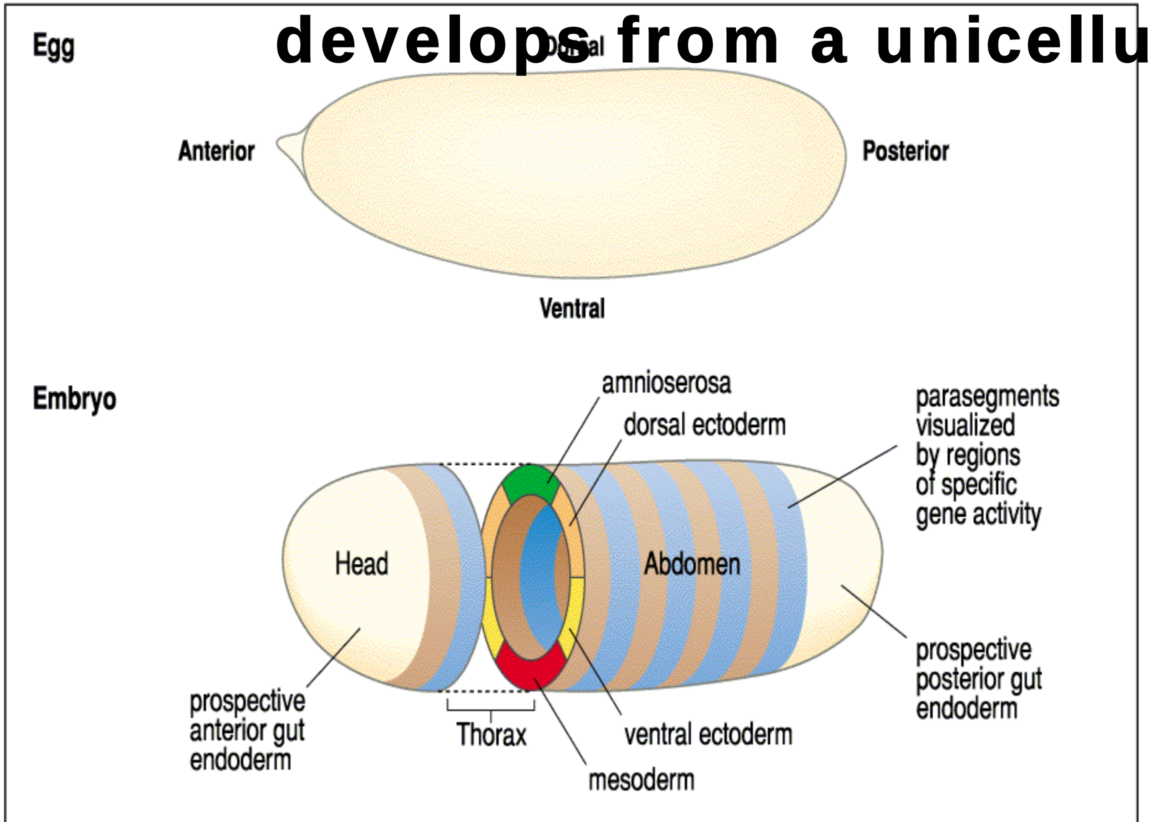
According to the widely accepted theory of natural selection, the whole process is blind. Genetic variations occur randomly and their selection by nature is purely based on their adaptability in given time and space.

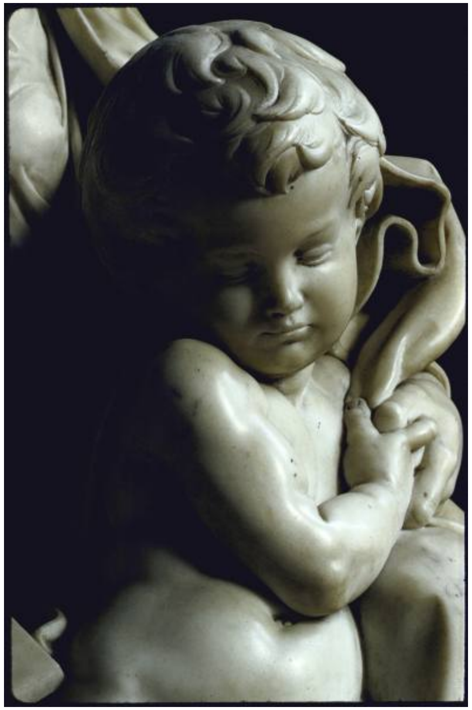
Evolution means *change*

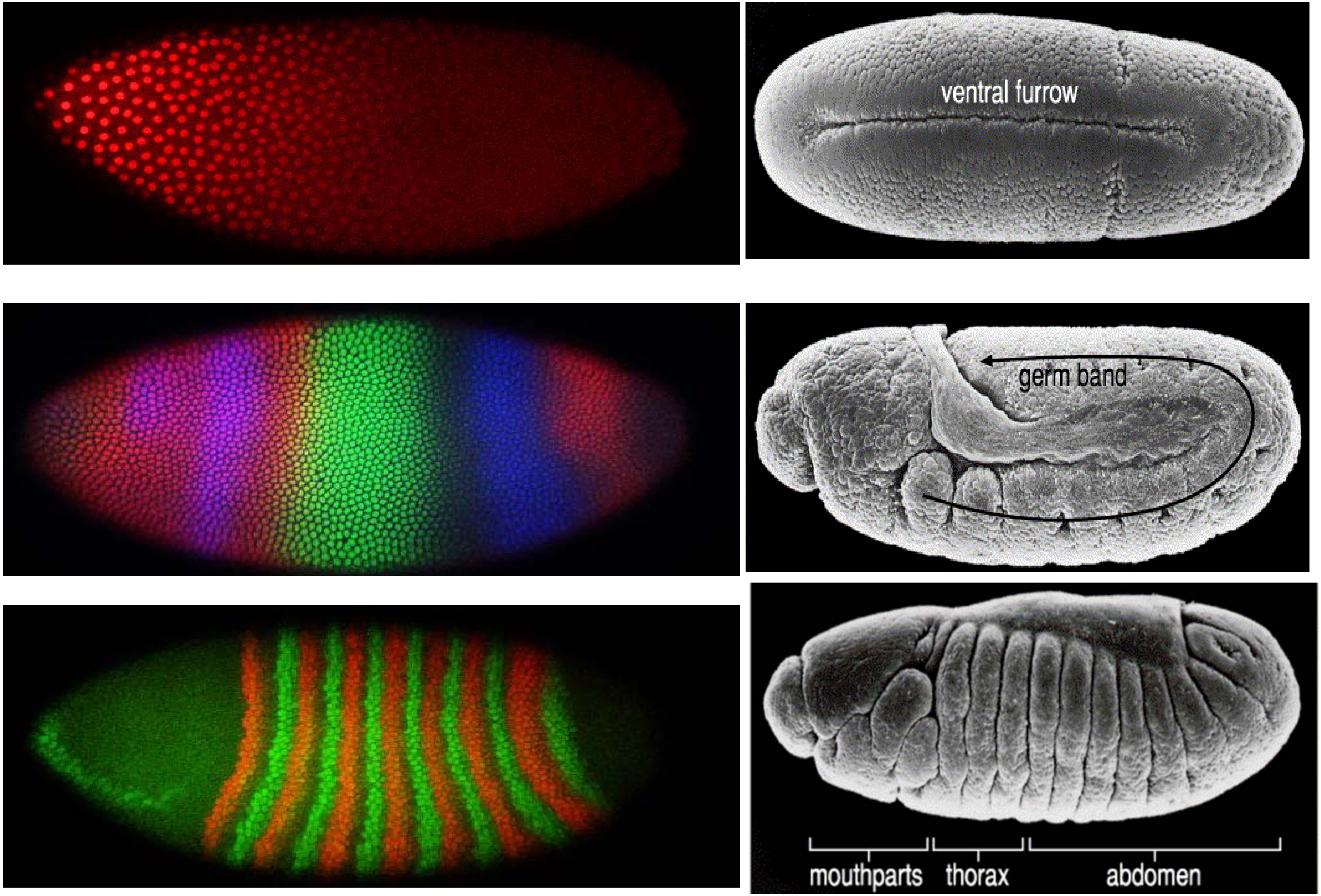
- **Evolution does not mean progress or improvement.**

- **It is a process of adaptation to survive in a constantly changing environmental condition.**

During development a multicellular organism develops from a unicellular embryo

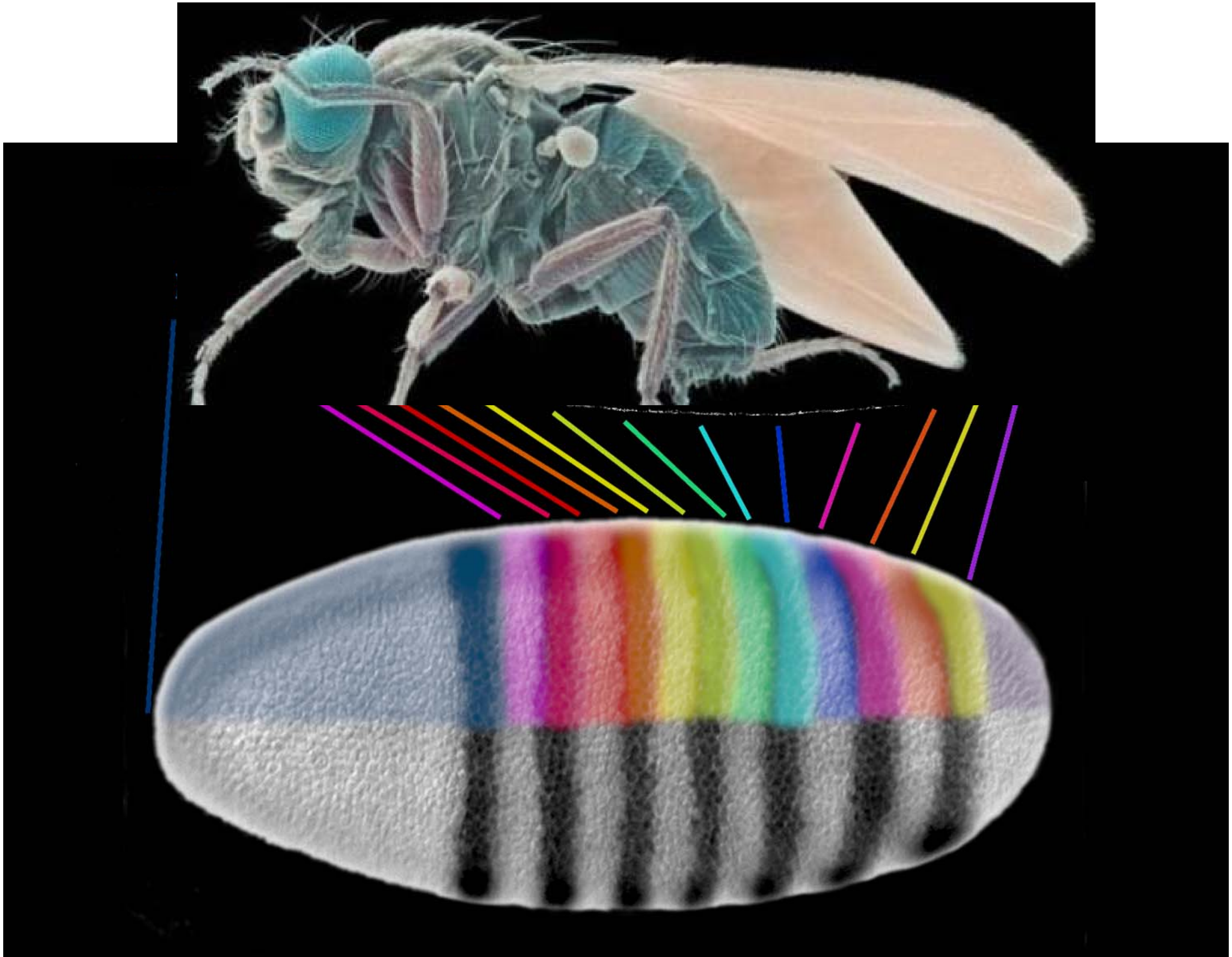






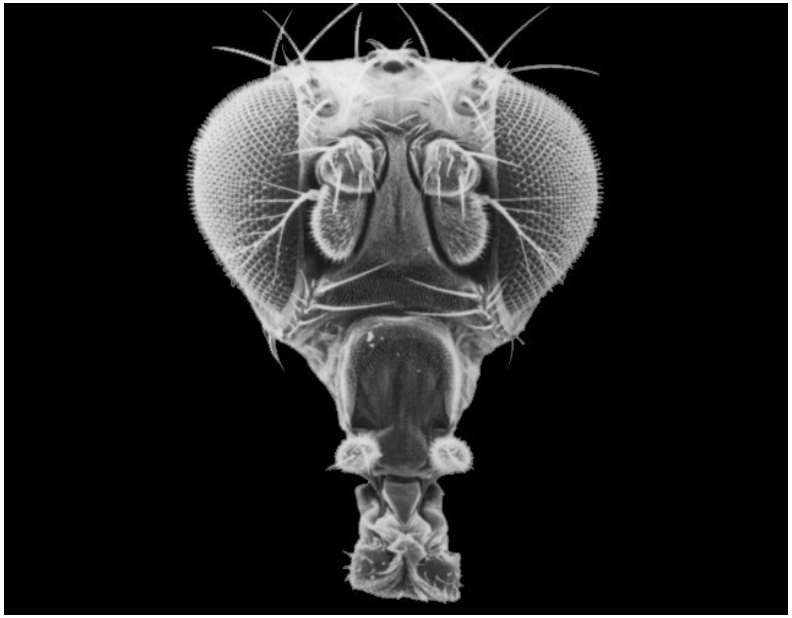
Morphological events are preceded by molecular events

Hox genes regulate segment specific developmental pathways

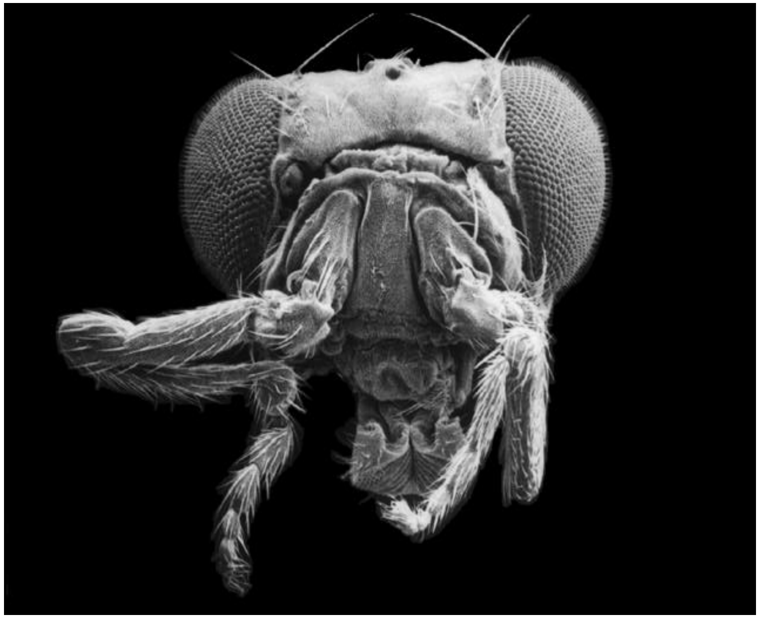


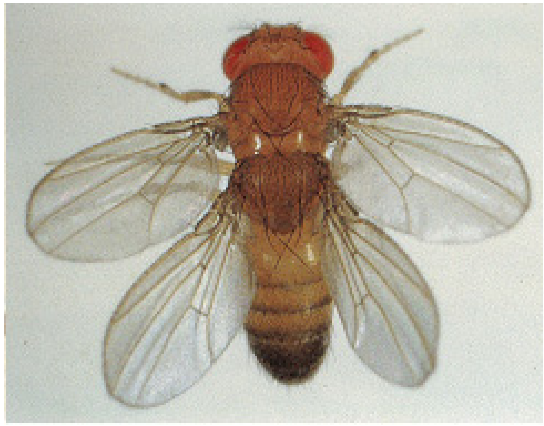
Hox genes specify body plan

Wild Type

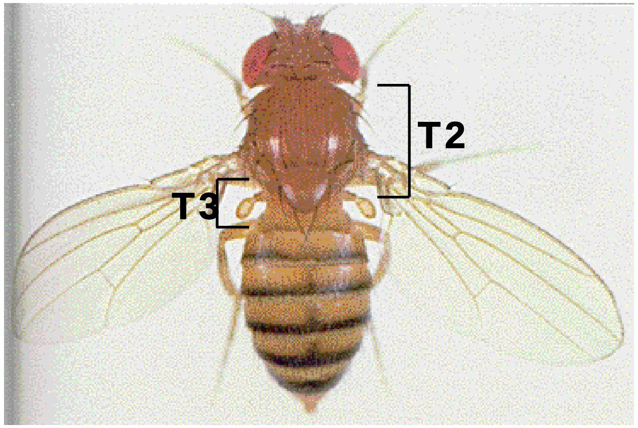


***Antp* Mutant**





No *Ubx* in T3

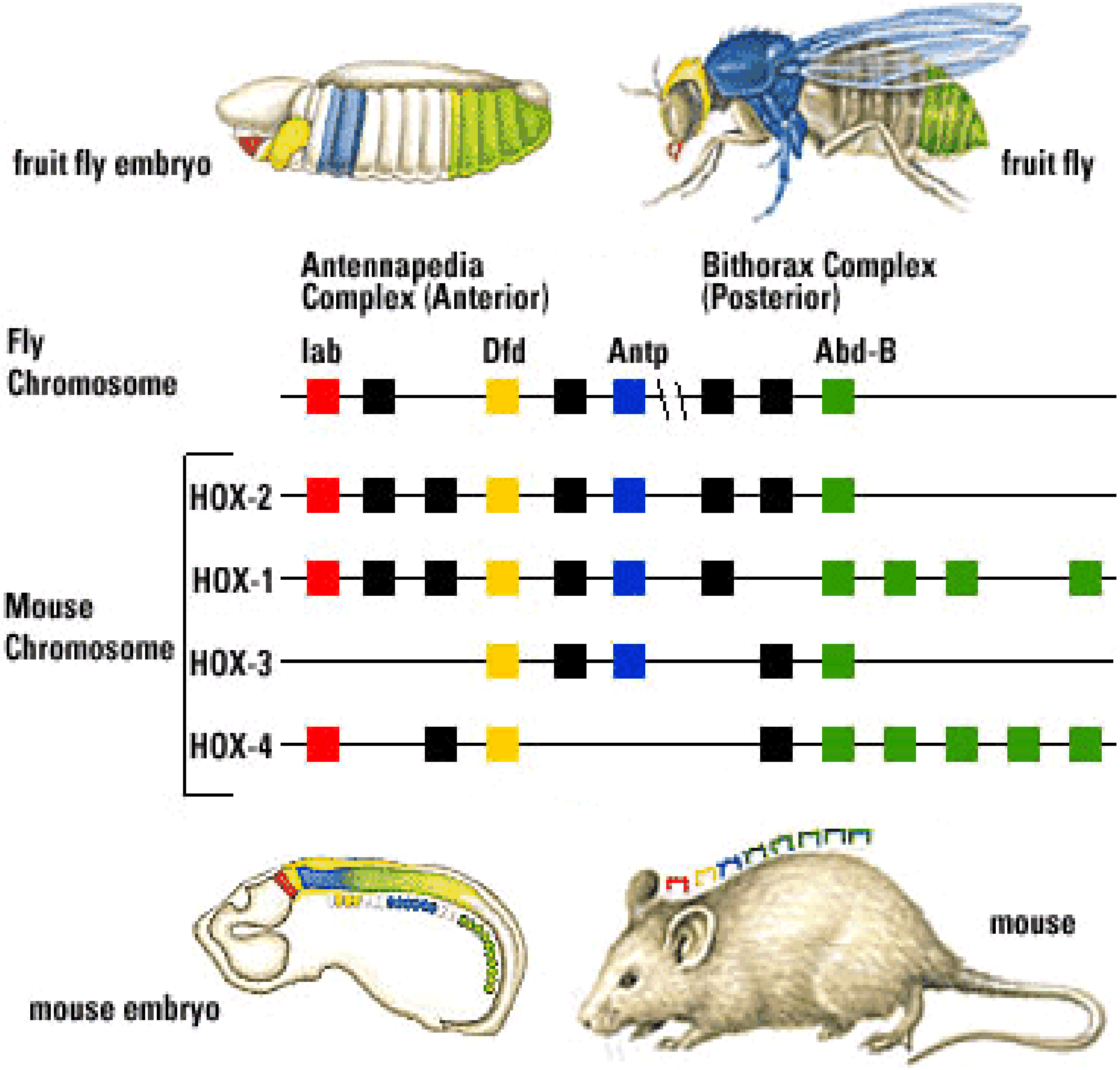


Wildtype

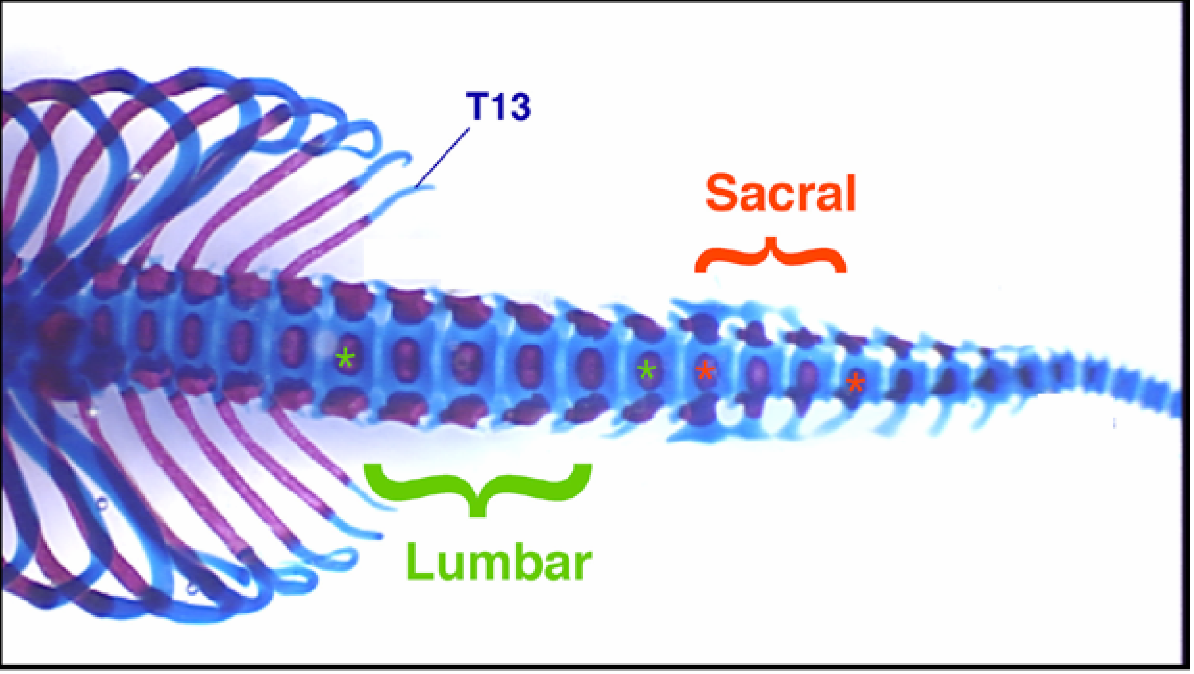
Ubx in both T2 and T3

Hox genes specify body plan

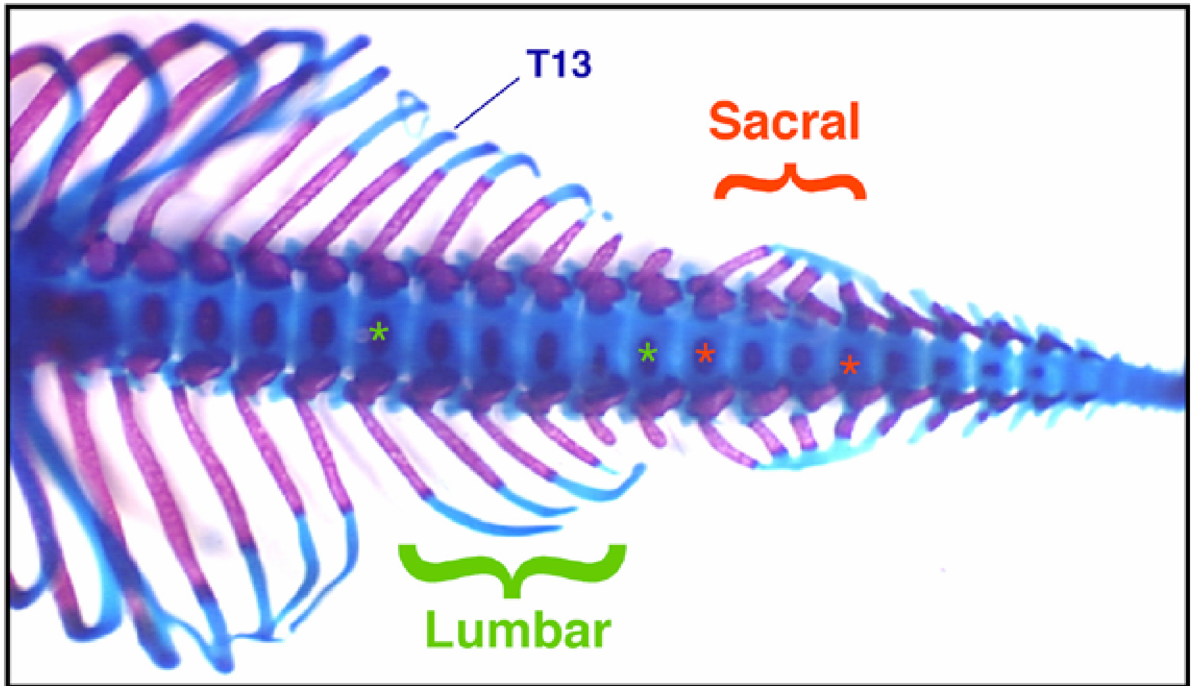
Comparison of Hox genes in fly and mouse embryos



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

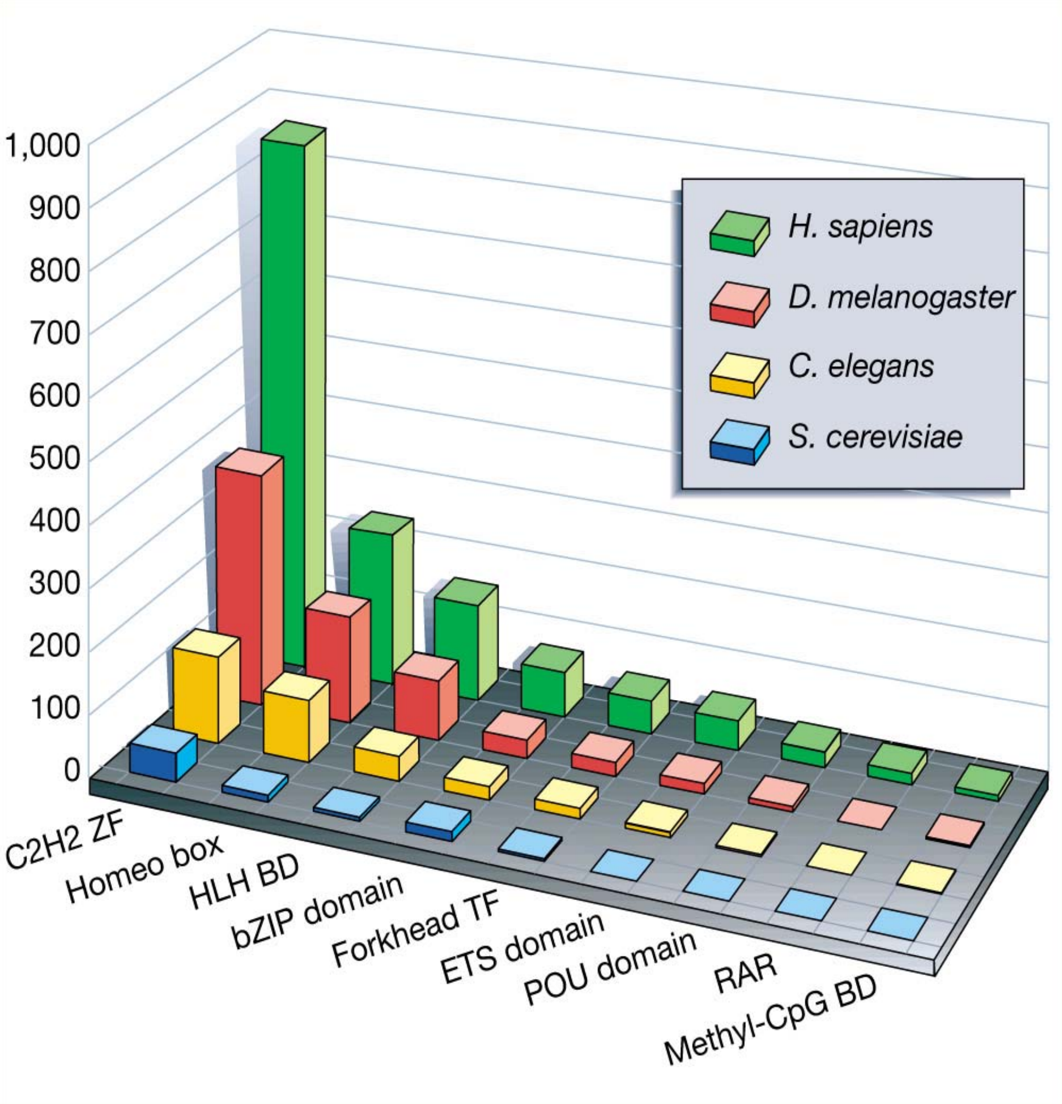
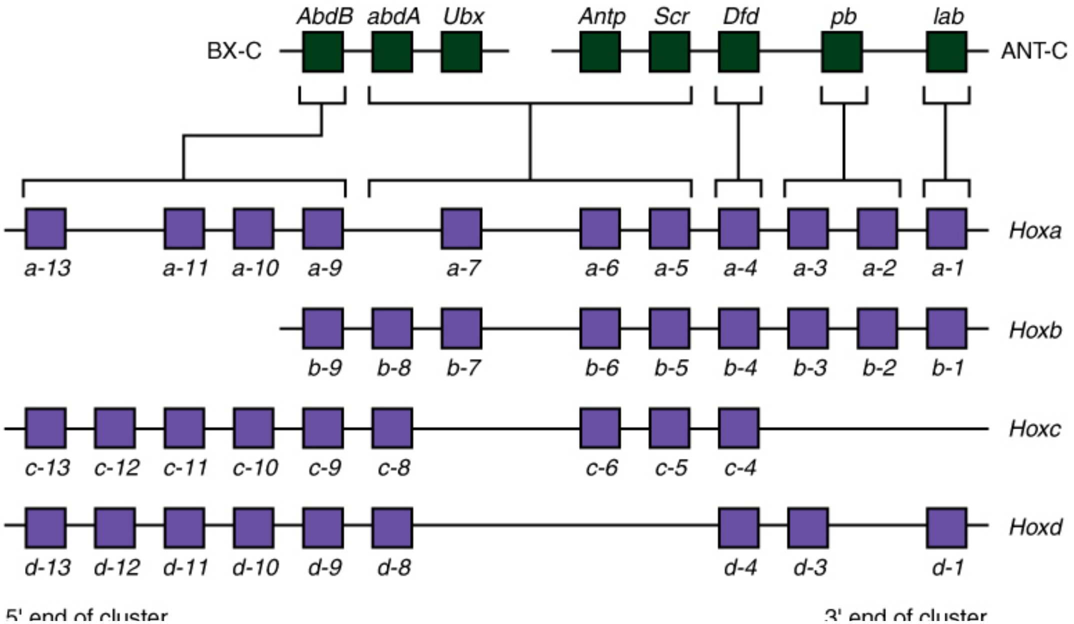
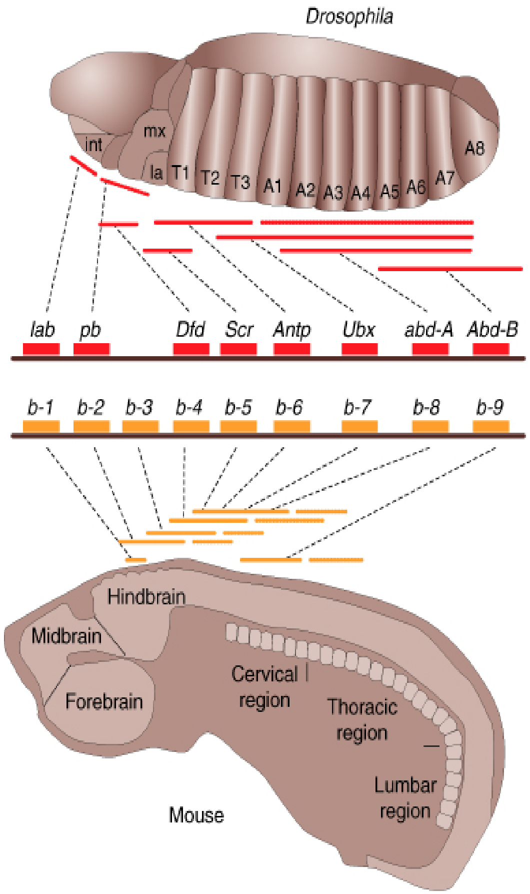


Figure 1 Genome-wide comparison of transcriptional activator families in eukaryotes. The relative sizes of transcriptional activator families among *Homo sapiens, D. melanogaster, C. elegans and S. cerevisiae* are indicated, derived from an analysis of eukaryotic proteomes using the INTERPRO database, 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

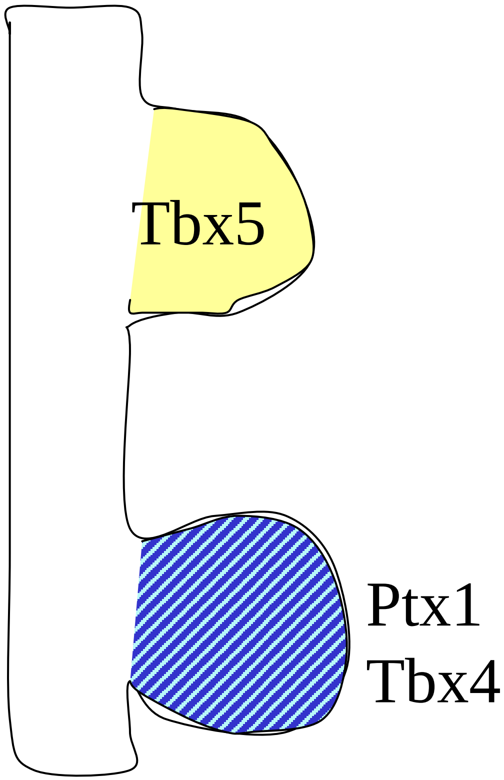




5’ end of cluster
expressed late in posterior body parts

3’ end of cluster
expressed early in anterior body parts

Fore Limb Vs Hind Limb



Hoxc4 Hoxc5

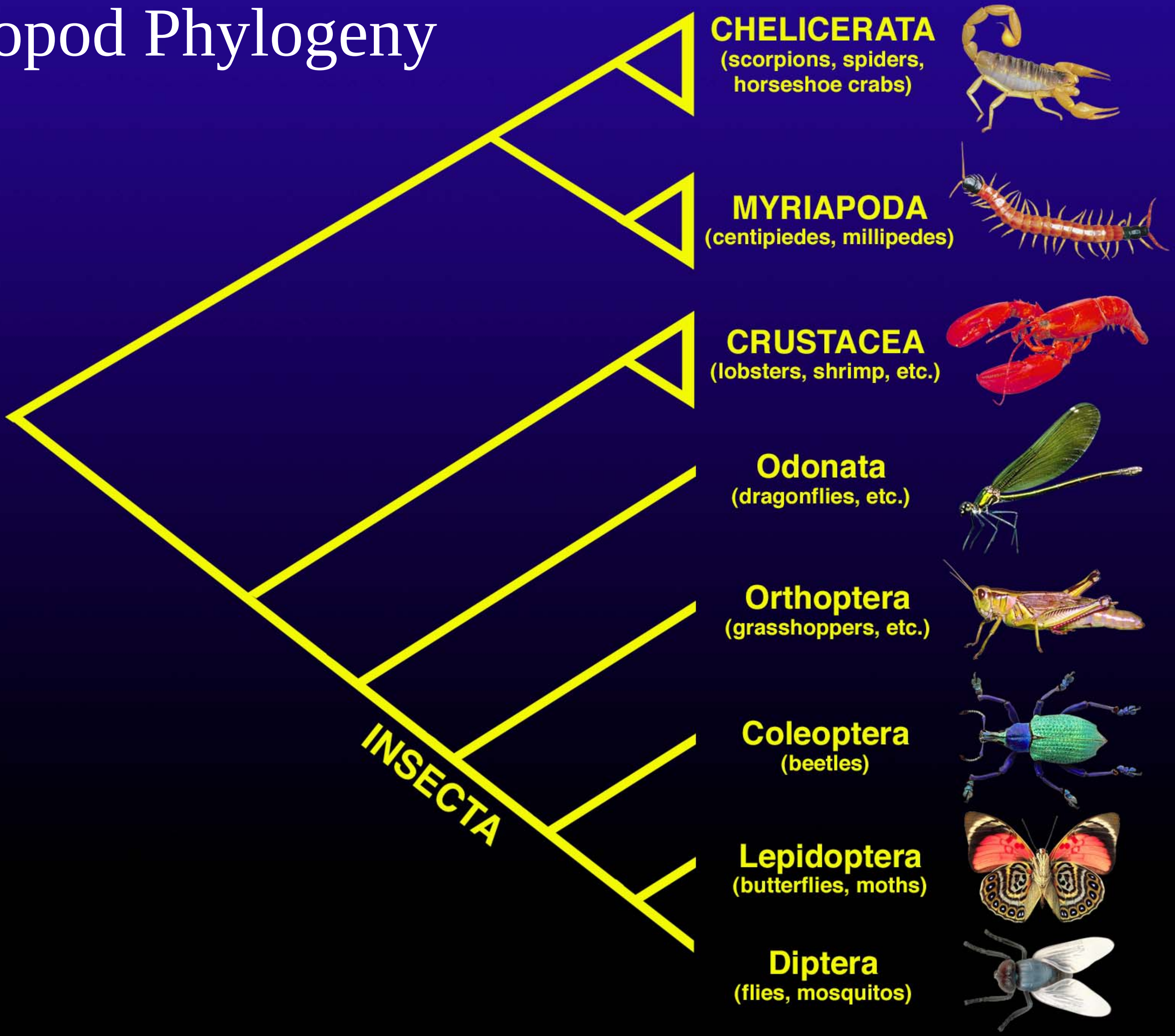
Hoxc6 Hoxc8

Hoxc9 Hoxc10 Hoxc11

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
(centipiedes, millipedes)

CRUSTACEA
(lobsters, shrimp, etc.)

Odonata
(dragonflies, etc.)

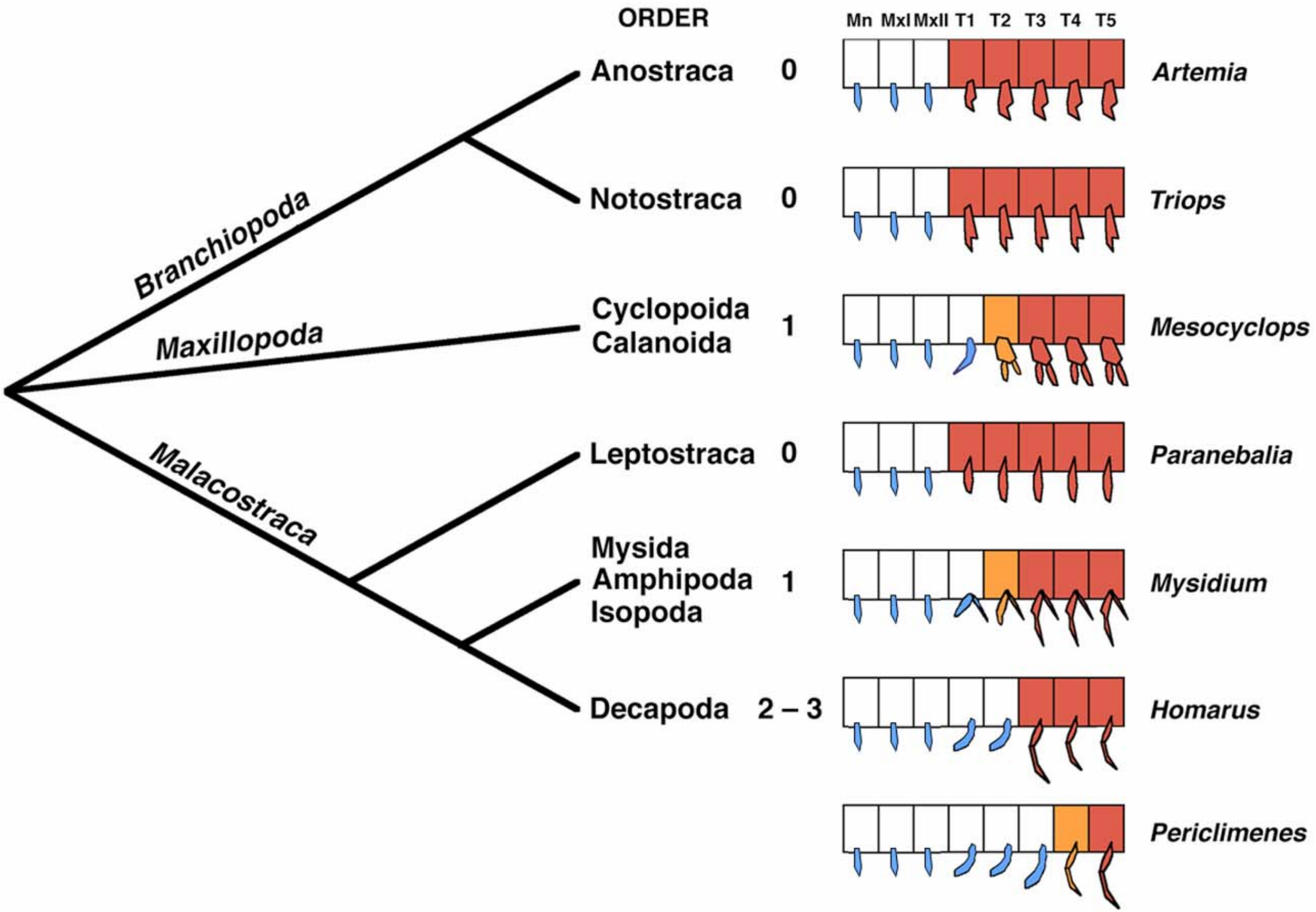
Orthoptera
(grasshoppers, etc.)

Coleoptera
(beetles)

Lepidoptera
(butterflies, moths)

Diptera
(flies, mosquitos)

INSECTA



Branchiopoda

Maxillopoda

Malacostraca

ORDER

Mn MxI MxII T1 T2 T3 T4 T5

Anostraca 0 *Artemia*

Notostraca 0 *Triops*

Cyclopoida
Calanoida 1 *Mesocyclops*

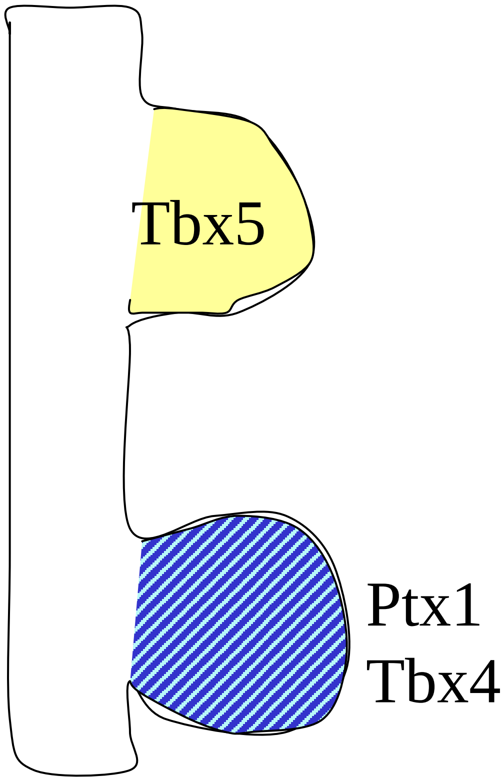
Leptostraca 0 *Paranebalia*

Mysida
Amphipoda 1 *Mysidium*
Isopoda

Decapoda 2 – 3 *Homarus*

Periclimenes

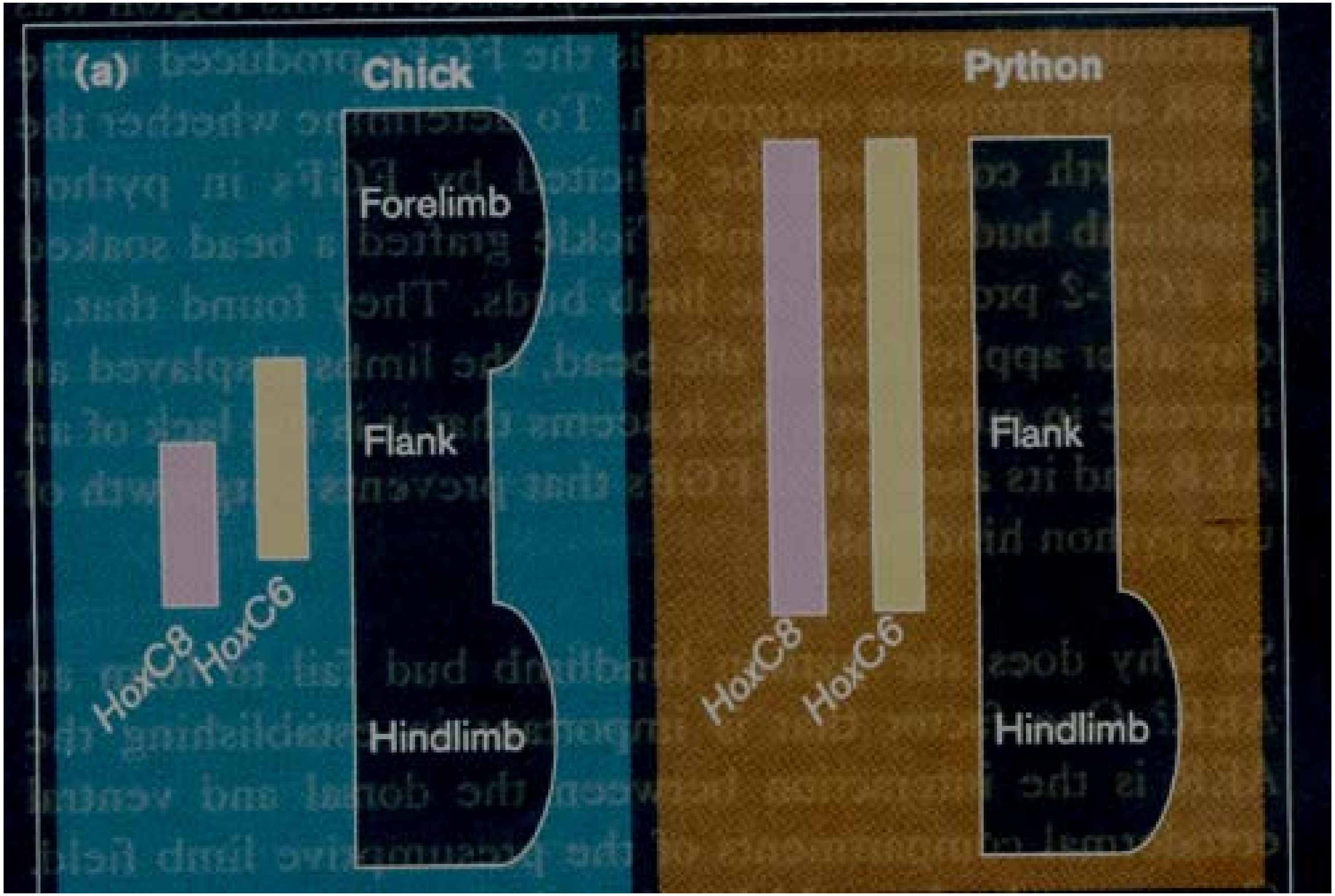
Fore Limb Vs Hind Limb



Hoxc4 Hoxc5

Hoxc6 Hoxc8

Hoxc9 Hoxc10 Hoxc11


(a)
Chick
Forelimb
Flank
Hindlimb
HoxC8
HoxC6
Python
Flank
Hindlimb
HoxC8
HoxC6

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

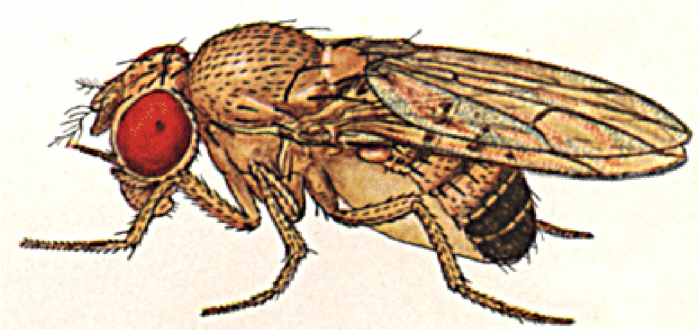
Class Onycophora

Common brine shrimp (*Artemia*)
Class Crustacea

Butterfly

Class Insecta

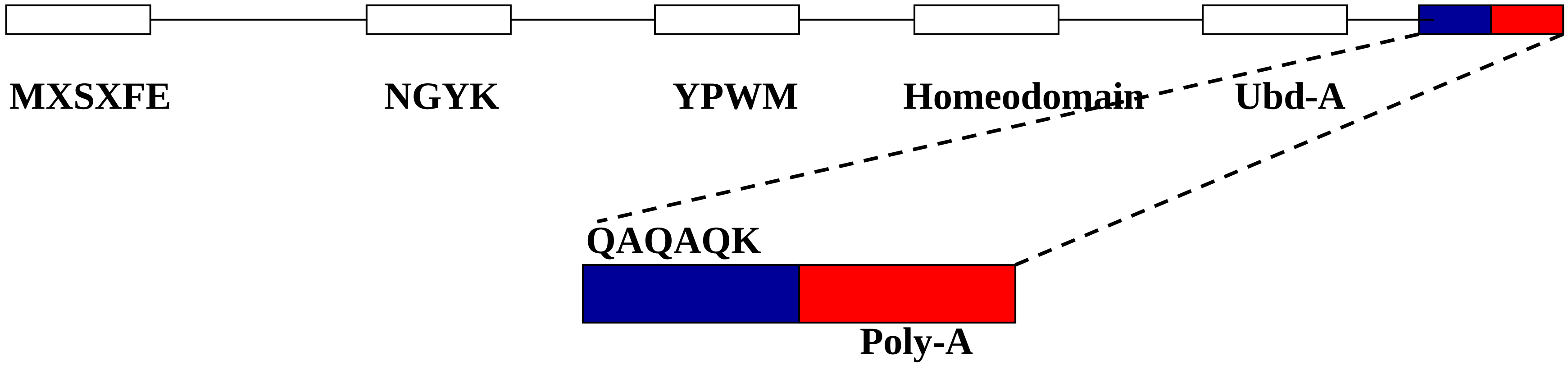
Order lepidoptera



Drosophila

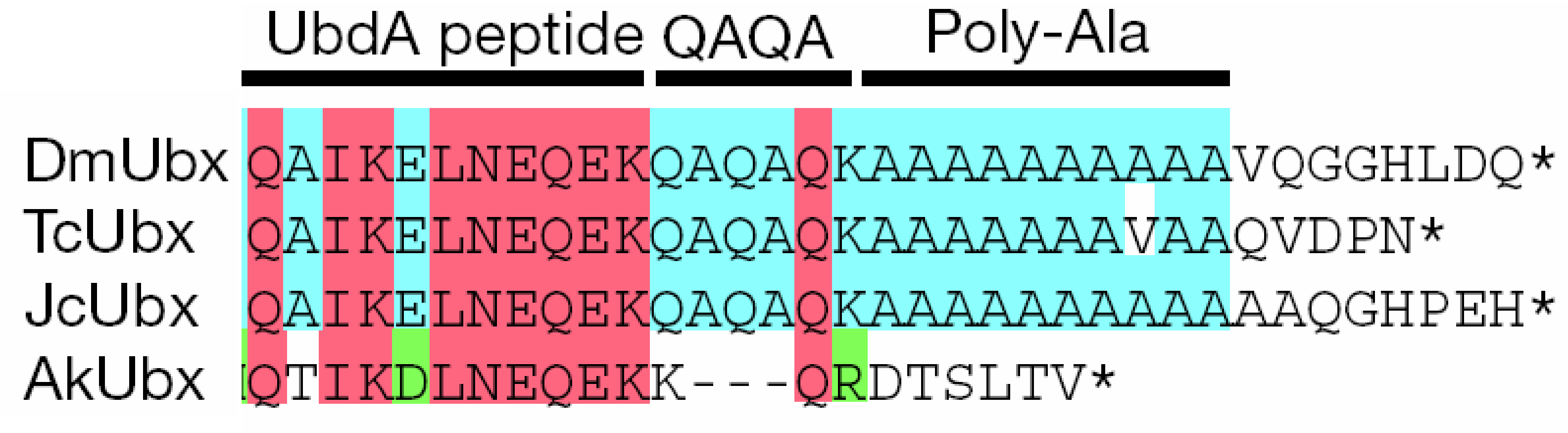
Class Insecta

Order Diptera



```
DmUbx TKSKIRSDLTQYGGISTDMGKRYSESLAGSLLPDWLGTNGL-RRRGRQTYTRYQTLELEKEFHTNHYLTR
TcUbx ------------------------------------GANGL-RRRGRQTYTRYQTLELEKEFHTNHYLTR
JcUbx ------------------------------------GANGL-RRRGRQTYTRYQTLELEKEFHTNHYLTR
AkUbx ------------------------------------GANGLQRKRGRQTYTRYQTLELEKEFHTNHYLTR

DmUbx RRRIEMAHALCLTERQIKIWFQNRRMKLKKEIQAIKELNEQEKQAQAQKAAAAAAAAAAVQGGHLDQ*
TcUbx RRRIEMAHALCLTERQIKIWFQNRRMKLKKEIQAIKELNEQEKQAQAQKAAAAAAAVAAQVDPN*
JcUbx RRRIEMAHALCLTERQIKIWFQNRRMKLKKEIQAIKELNEQEKQAQAQKAAAAAAAAAAAAQGHPEH*
AkUbx RRRIEMAHALCLTERQIKIWFQNRRMKLKKEMQTIKDLNEQEKK---QRDTSLTV*
```

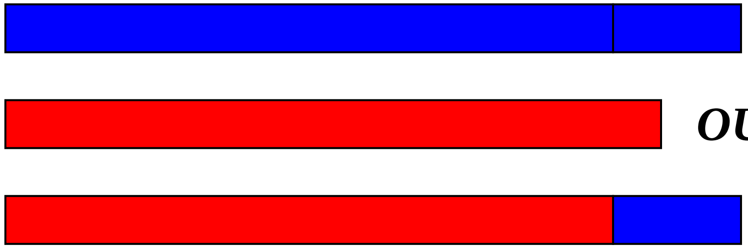


```
DmUbx QAIKELNEQEKQAQAQKAAAAAAAAAAAVQGGHLDQ*
TcUbx QAIKELNEQEKQAQAQKAAAAAAAAVAAQVDPN*
JcUbx QAIKELNEQEKQAQAQKAAAAAAAAAAAAAQGHPEH*
AkUbx QTIKDLNEQEKK---QRDTSLTV*
```

Poly-alanine rich and glutamine/alanine rich sequences found in many repression domains.

Mediate repression by interacting with basal transcriptional machinery.

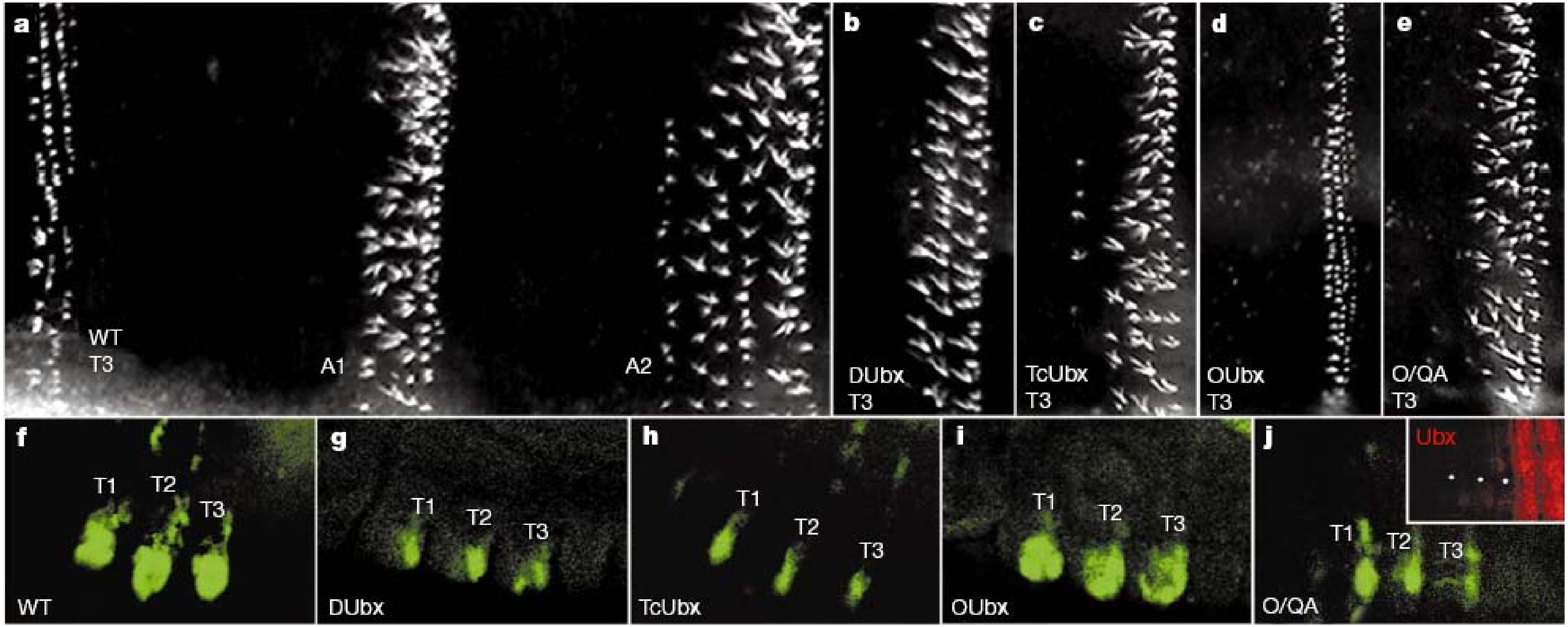
N terminal



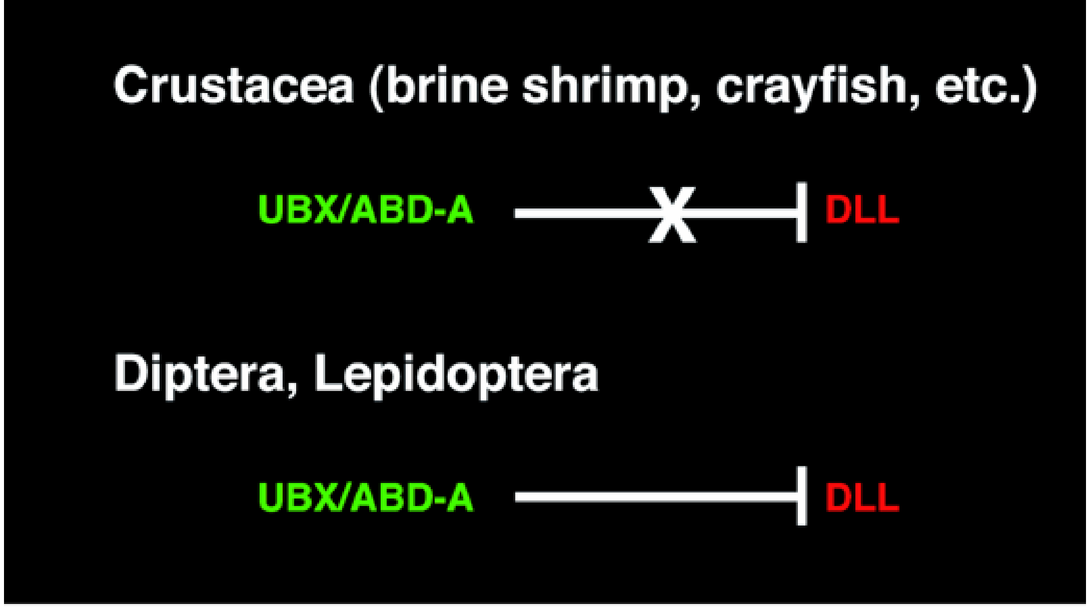
DUbx

OUbx

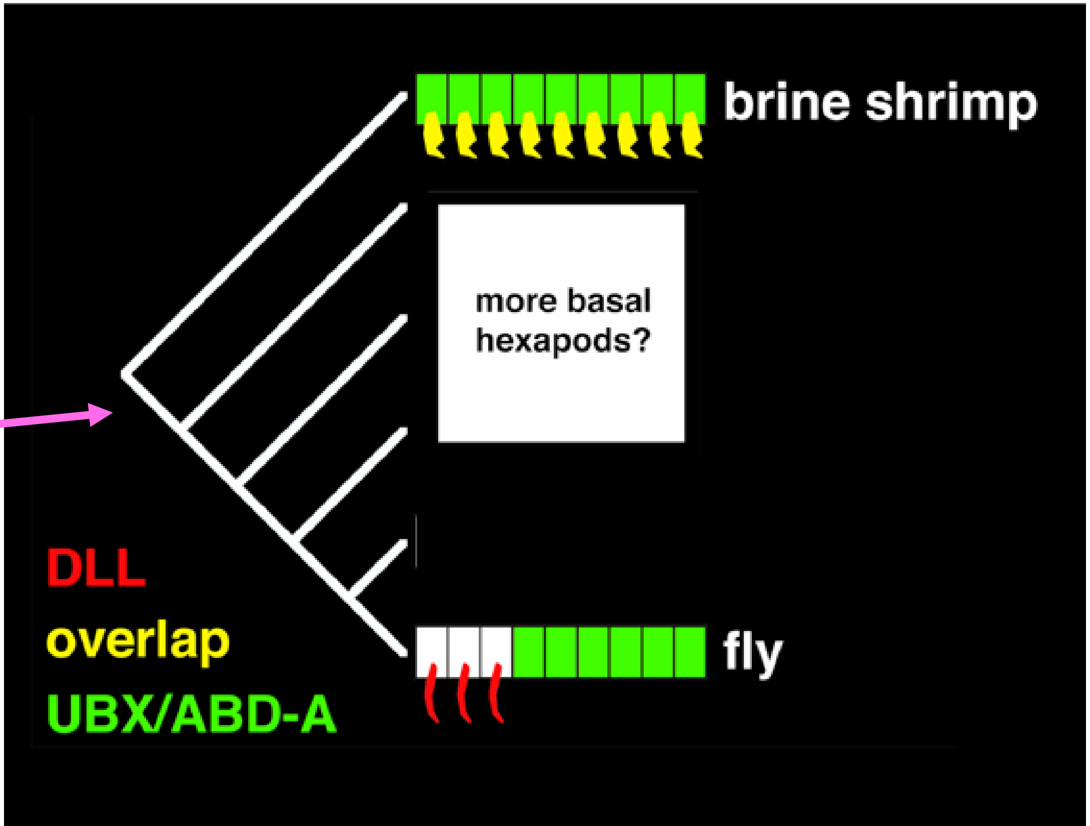
OUbx /QA



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

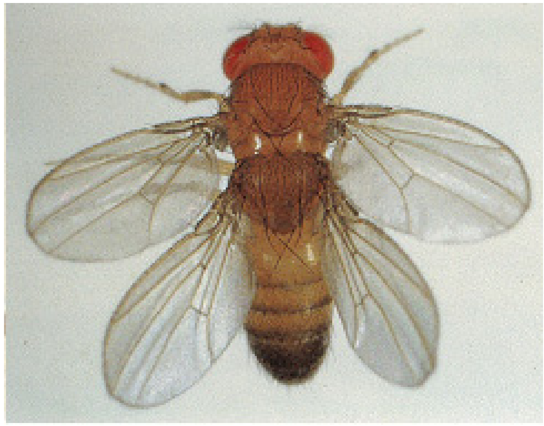
- **Changes in the number of Hox gene (duplication and divergence)**

- **Changes in domain of Hox gene expression**

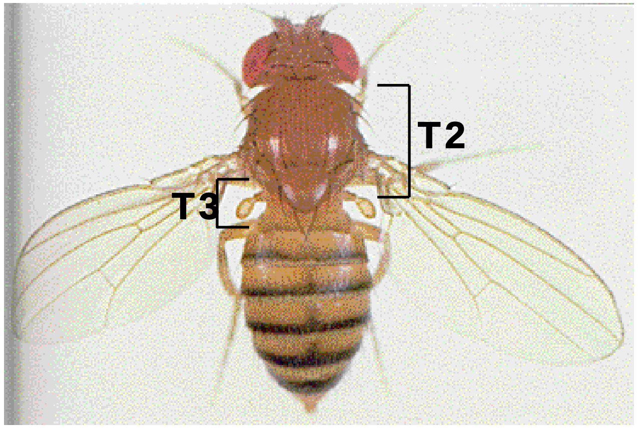
- **Changes in Hox gene that gives the protein new properties**

- **Changes in Hox-protein responsive elements of downstream genes**

Downstream of Homeotic genes…



No *Ubx* in T3



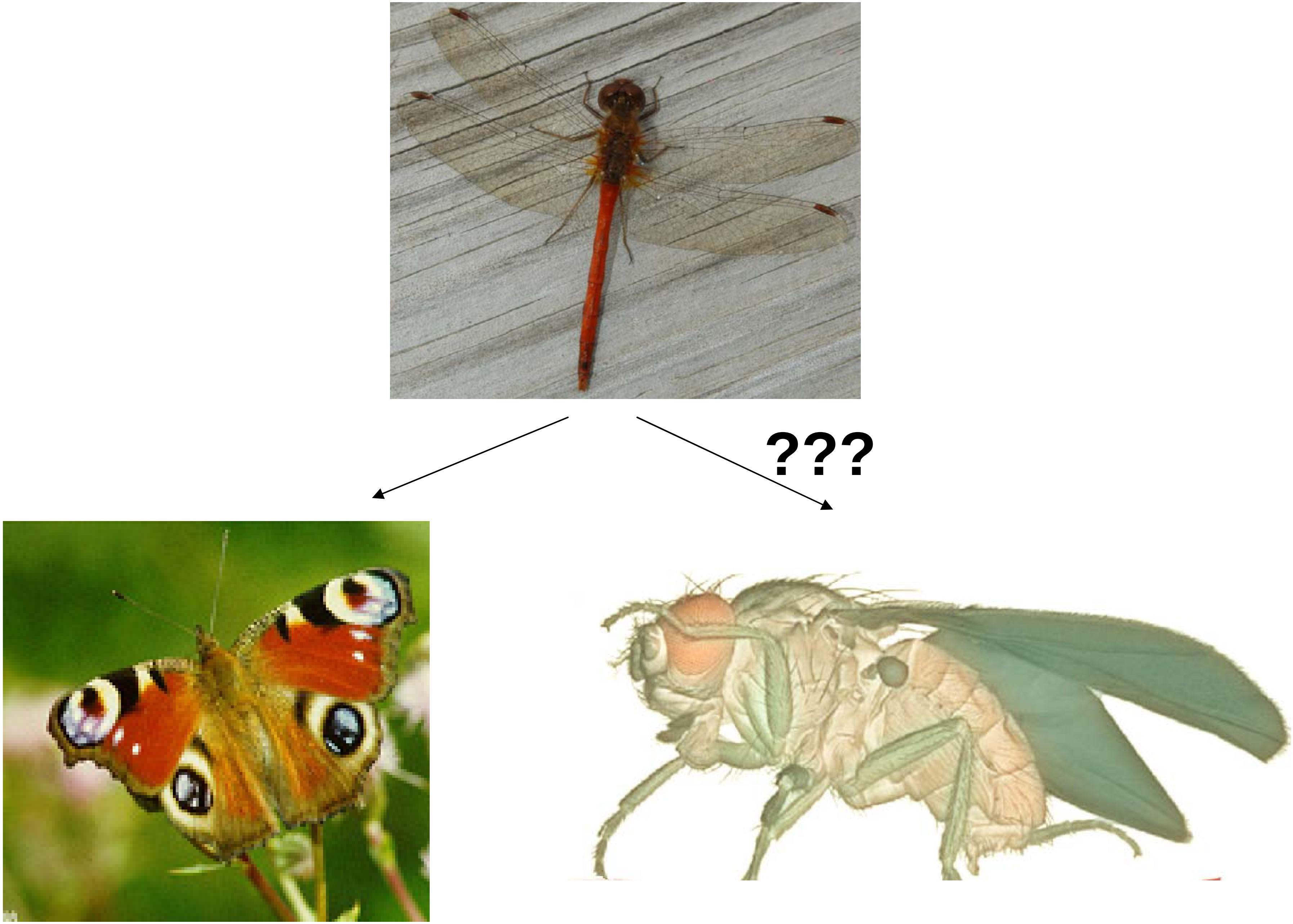
Wildtype

Ubx in both
T2 and T3

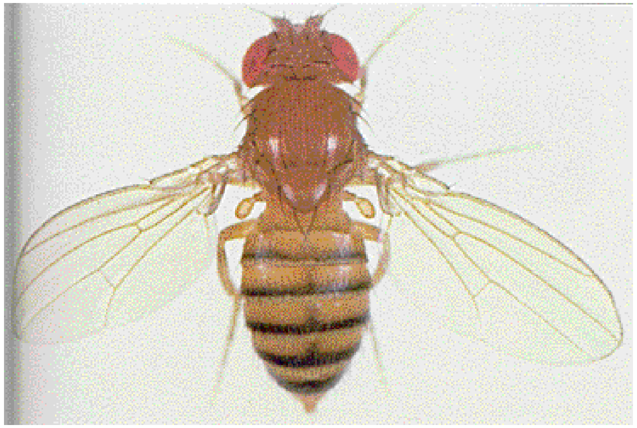
Organ identity: wing vs haltere

homeotic mutant

dragonfly



???



Wildtype

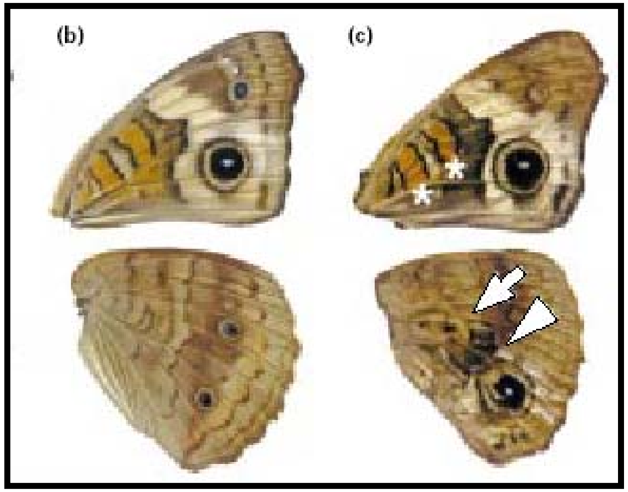
***Ubx* in both T2 and T3**

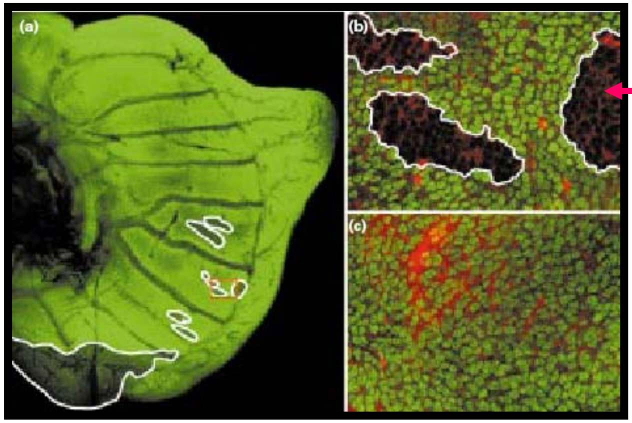
over-expression of Ubx from butterflies and even from a non-winged arthropod such as Onychophora is sufficient to induce wing-to-haltere transformations in *Drosophila*

Grenier, J. K., Carroll, S. B. 2000. Functional evolution of the Ultrabithorax protein. *Proc. Natl. Acad. Sci., USA* 97, 704–709.

The difference must lie in the response of the downstream targets of Ubx

In butterflies…



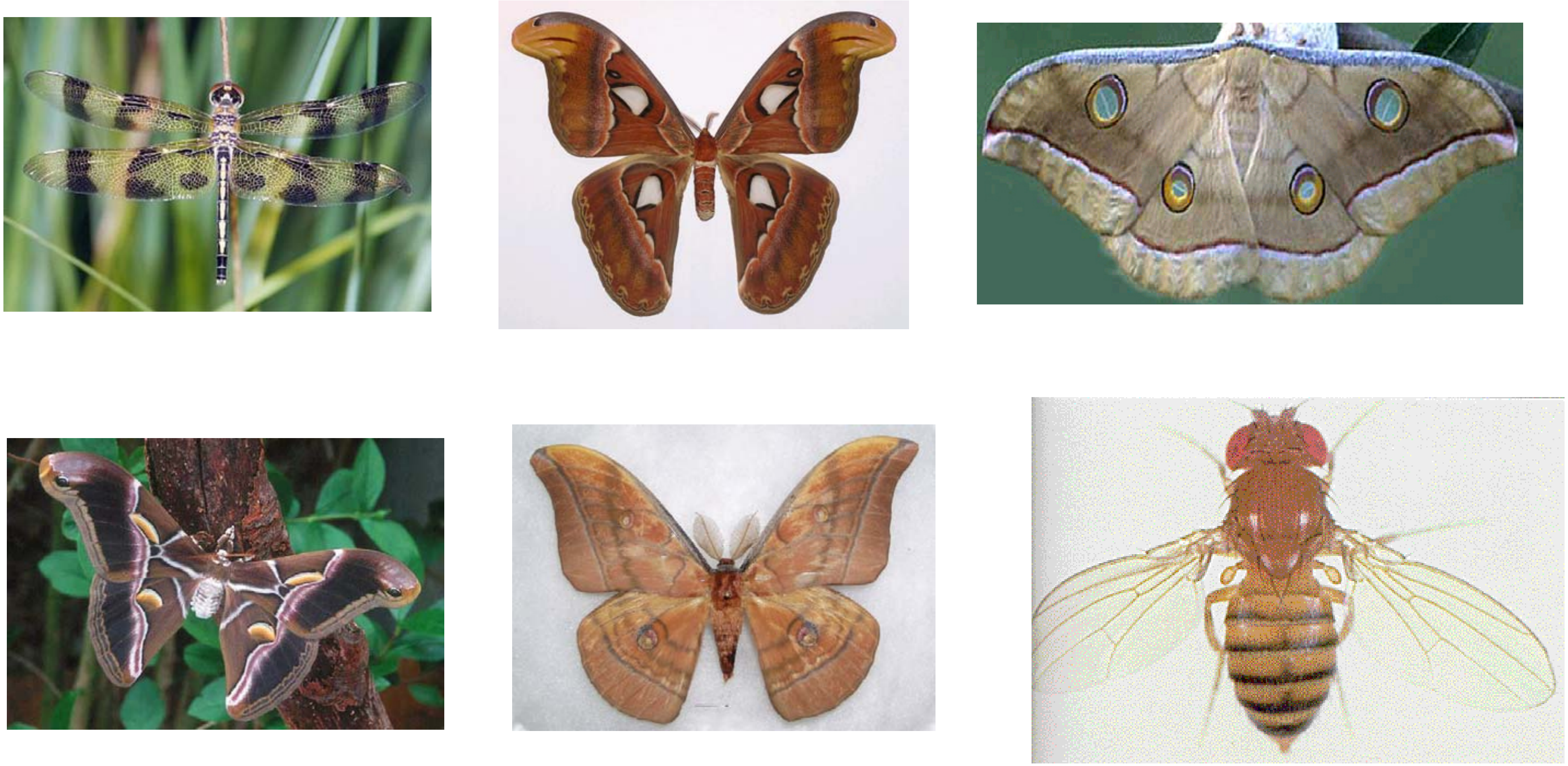


Loss of
Ubx

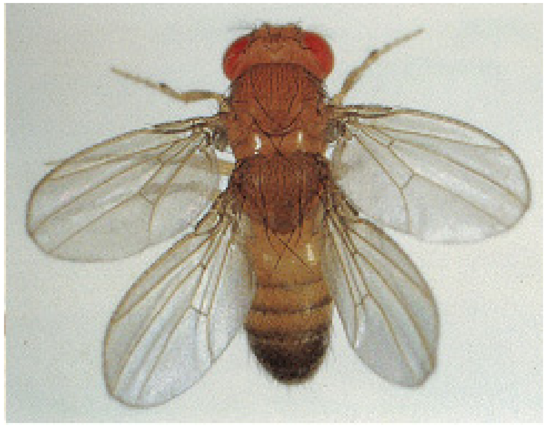
***Hind sight* mutants exhibit similarity between fore and hind wings**

Ubx is required for hind wing identity in butterflies

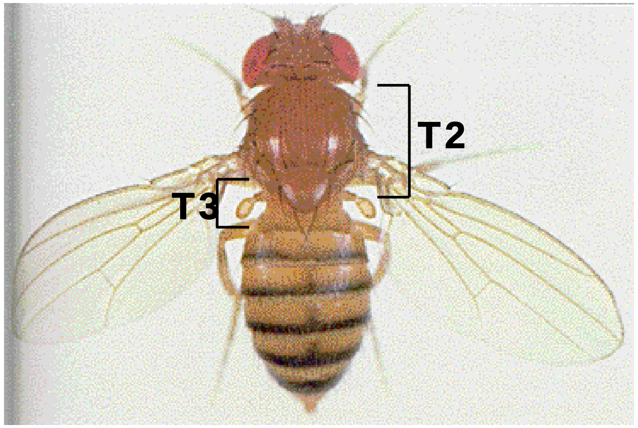
During dipteran evolution, certain wing-patterning genes must have come under the regulation of *Ubx*.



Downstream of Homeotic genes…



No *Ubx* in T3



Wildtype

Ubx in both
T2 and T3

Organ identity: wing vs haltere

Summary,

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

***Developmental Biology* 212, 491-502 (1999)**
***Development* 130, 1537-47 (2003)**
***Mechanism of Development (GEP)* 5, 113-121 (2004)**
***J Cell Science* 117, 1911-1922 (2004)**
***Development* 131, 1007-1016 (2004)**
***J Cell Science* 118, 3363-3370. (2005)**
***Developmental Biology* 291, 356-367 (2006)**
***Developmental Biology* 296, 340-352 (2006)**
***Developmental Biology* 302, 243-255 (2007)**

The difference must lie in the response of the downstream targets of Ubx

Several models linking Hox evolution to changes in adult body plan

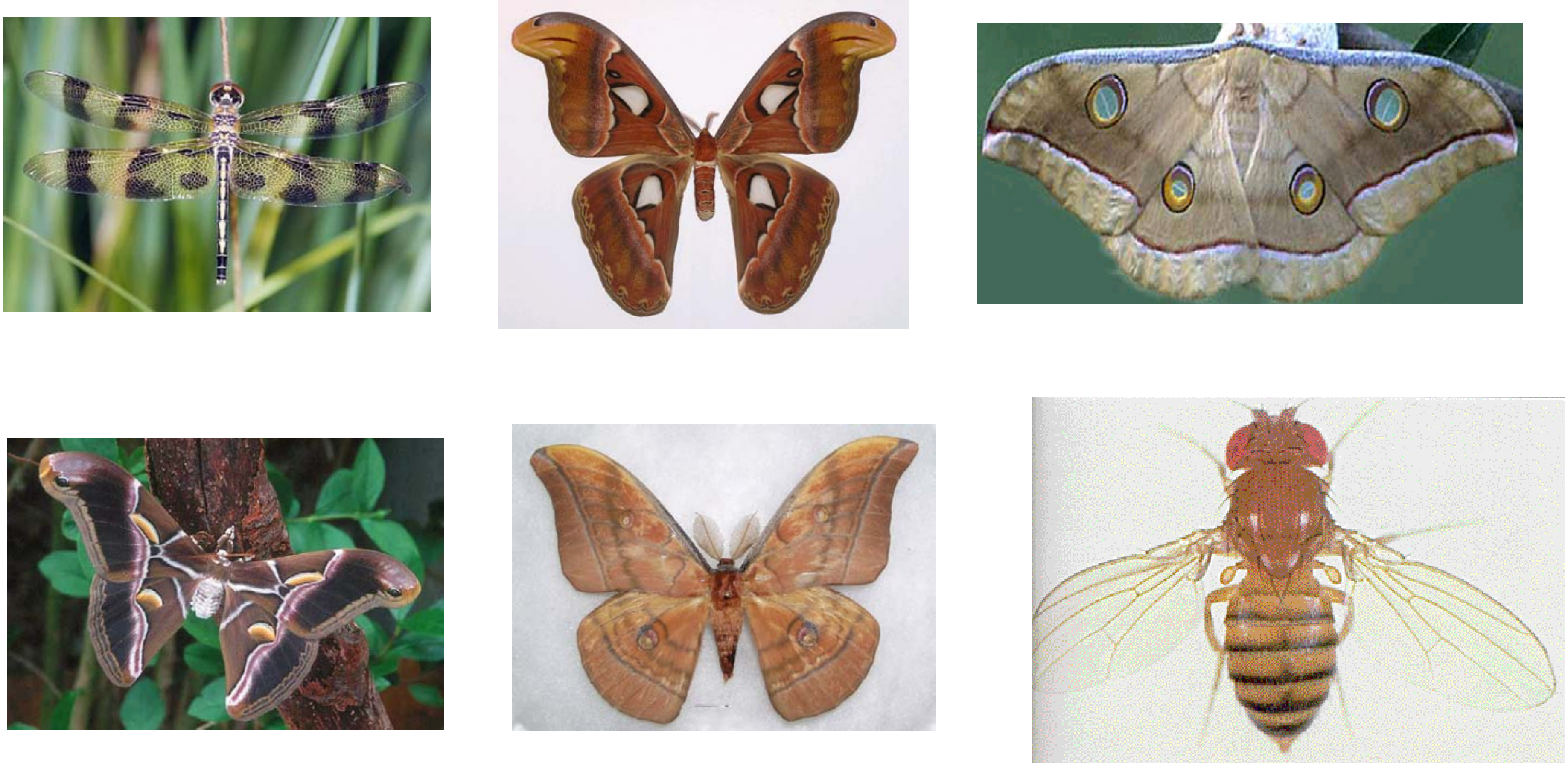
- **Changes in the number of Hox gene (duplication and divergence)**

- **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 Dipteran-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

Tribolium

Anopheles

Apis

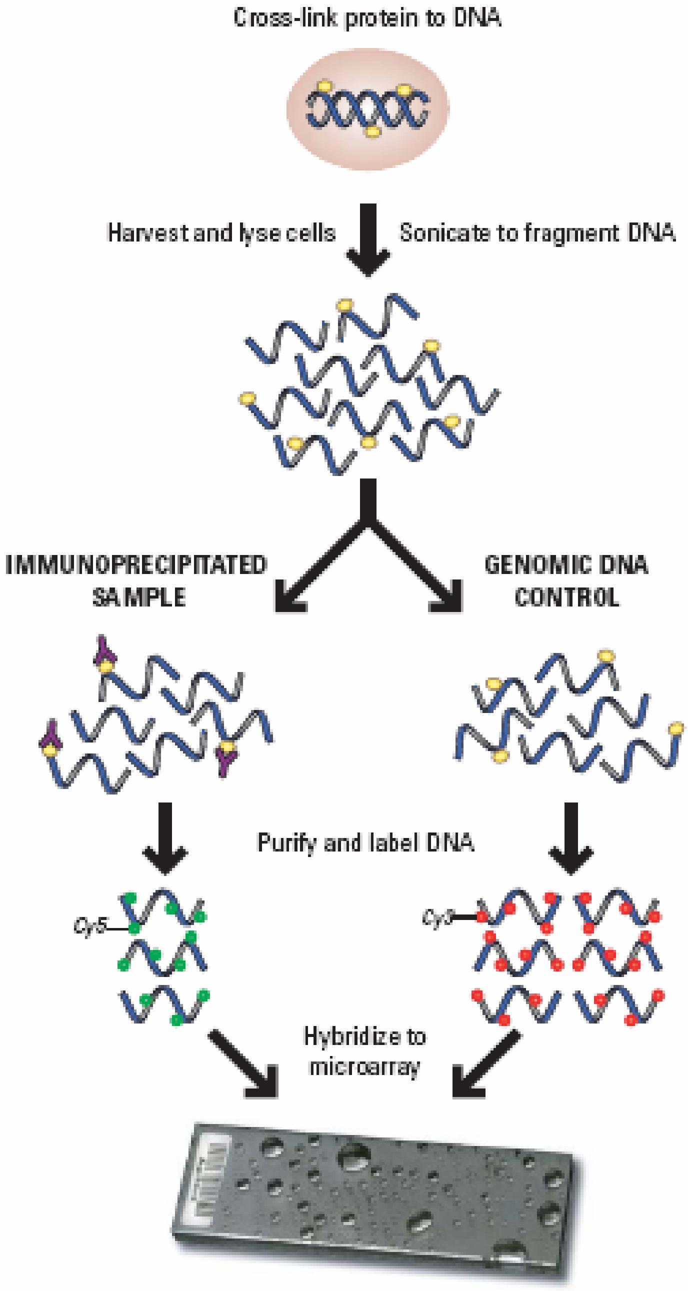
Bombyx

Drosophila

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



Cross-link protein to DNA

Harvest and lyse cells

Sonicate to fragment DNA

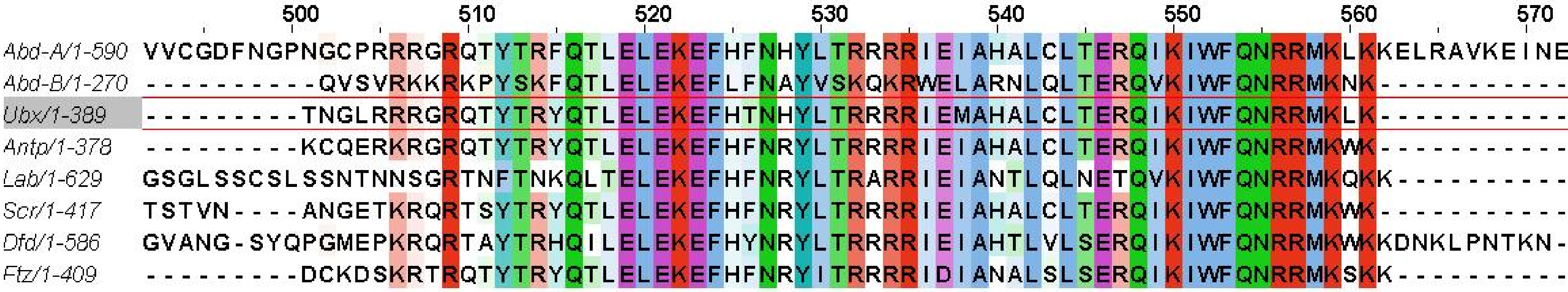
IMMUNOPRECIPITATED SAMPLE

GENOMIC DNA CONTROL

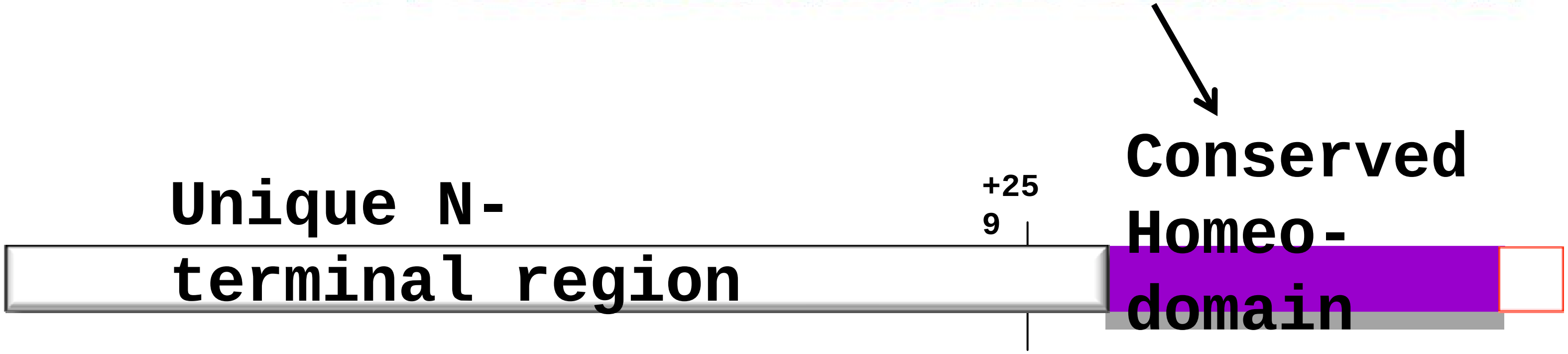
Purify and label DNA

Hybridize to microarray

•Polyclonal Antibodies generated against N-terminal region of Drosophila Ubx



```
                     500       510       520       530       540       550       560       570
Abd-A/1-590   VVCGDFNGPNGCPRRRGRQTYTRFQTLELEKEFHFNHYLTRRRRIEIAHALCLTERQIKIWFQNRRMKLKKELRAVKEINE
Abd-B/1-270   ----------QVSVRKKRKPYSKFQTLELEKEFLFNAYVSKQKRWELARNLQLTERQVKIWFQNRRMKNK-----------
Ubx/1-389     ---------TNGLRRRGRQTYTRYQTLELEKEFHTNHYLTRRRRIEMAHALCLTERQIKIWFQNRRMKLK-----------
Antp/1-378    ---------KCQERKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWK-----------
Lab/1-629     GSGLSSCSLSSNTNNSGRTNFTNKQLTELEKEFHFNRYLTRARRIEIANTLQLNETQVKIWFQNRRMKQKK----------
Scr/1-417     TSTVN----ANGETKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWK-----------
Dfd/1-586     GVANG-SYQPGMEPKRQRTAYTRHQILELEKEFHYNRYLTRRRRIEIAHTLVLSERQIKIWFQNRRMKWKKDNKLPNTKN-
Ftz/1-409     ---------DCKDSKRTRQTYTRYQTLELEKEFHFNRYITRRRRIDIANALSLSERQIKIWFQNRRMKSKK----------
```



Unique N-terminal region

+259

Conserved Homeo-domain

Post ChIP-chip

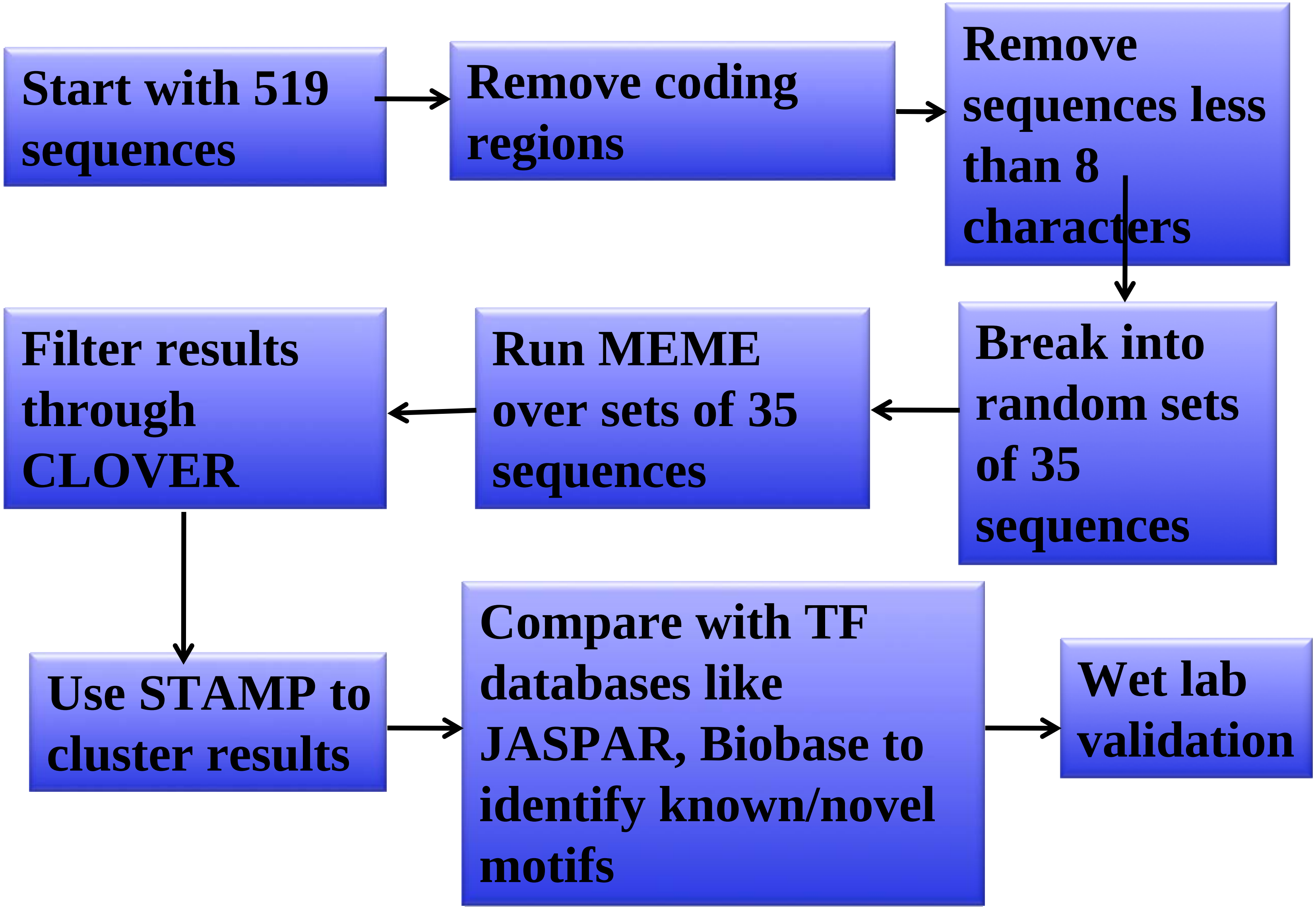
1. Validation by RNA in situ, q-PCR (independent of Ubx polyclonal antibodies)

2. Functional characterization of some interesting candidate genes

3. Data Mining

Overview of Analysis Strategy to find Motifs

T-T-A-A-T-T/G-A/G



Start with 519 sequences

Remove coding regions

Remove sequences less than 8 characters

Break into random sets of 35 sequences

Run MEME over sets of 35 sequences

Filter results through CLOVER

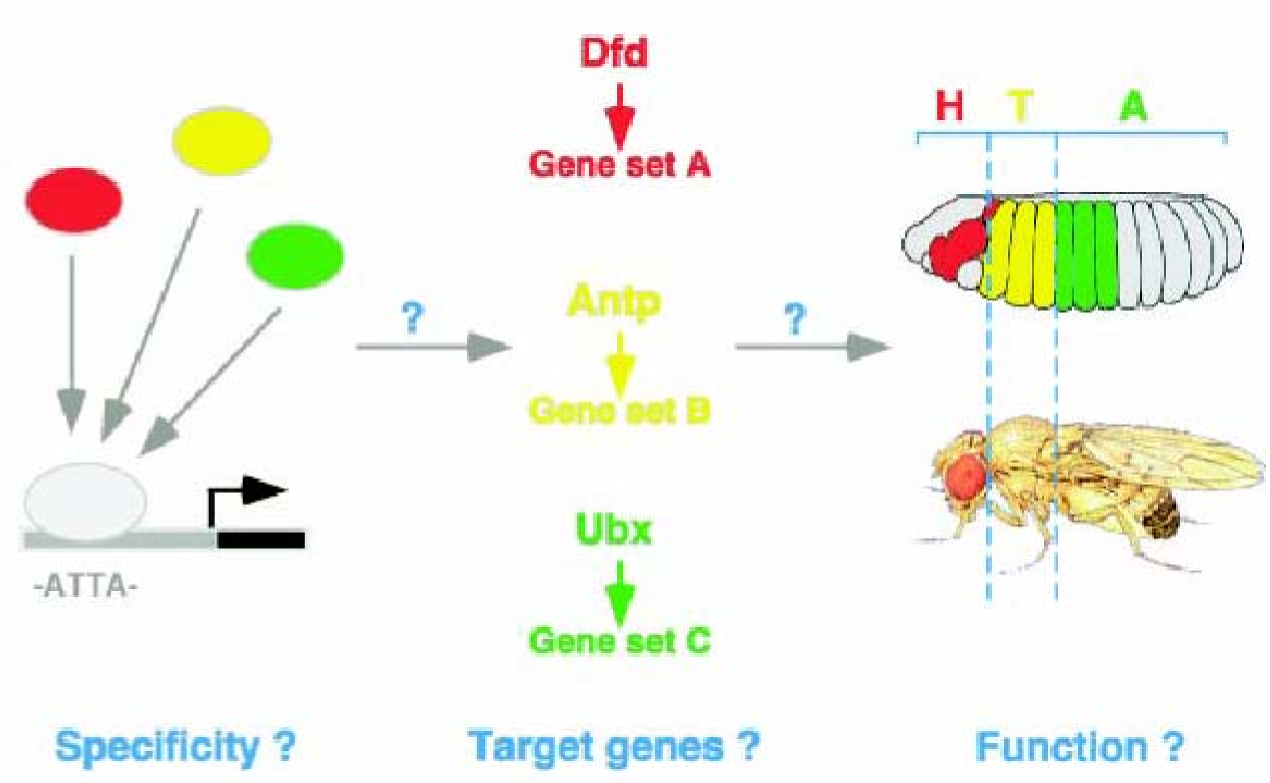
Use STAMP to cluster results

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

Wet lab validation

Hox Paradox

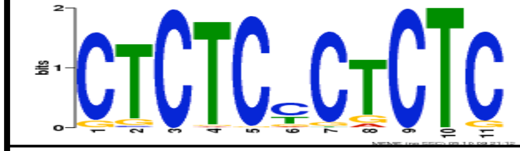
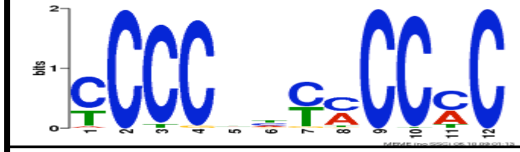
Ubx core binding sequence TAAT is a common binding site for many other Hox



Lohmann I. *et al*, 2008

How Hox proteins select specific targets in vivo?

Motifs found from probes with p<0.01 (255 Probes)

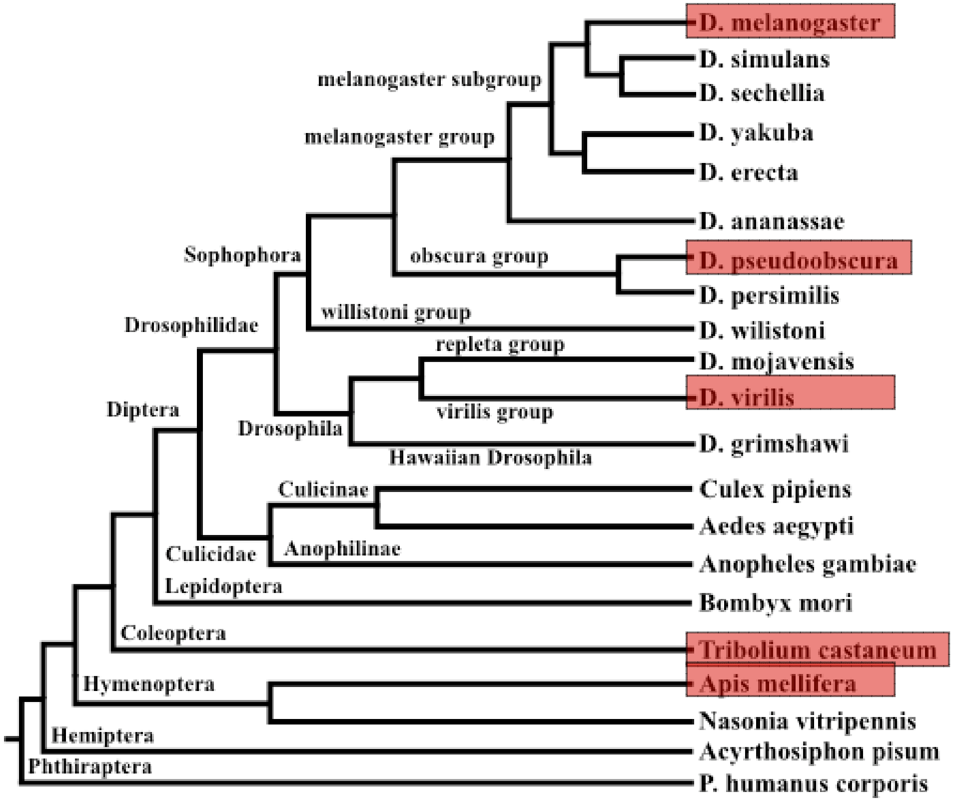
Motif	#Occur (%)	Consensus
	78(30.5)	**ACAC[AG]C[AG]CACAC**
	64(25.1)	**CTCTC[CT]CTCTC**
	82(32.2)	**[GA]CA[AG]CAACAACA**
	63(24.7)	**GCAG[AC][GC]GCAGC**
	66(25.9)	**[CT]CCC[TAC][CTA][CT] [CA]CC[CA]C**
	51(20.0)	**G[ACG]GA[GA][AC]GAG[CA]G[CA]**

Data from ChIP-chip

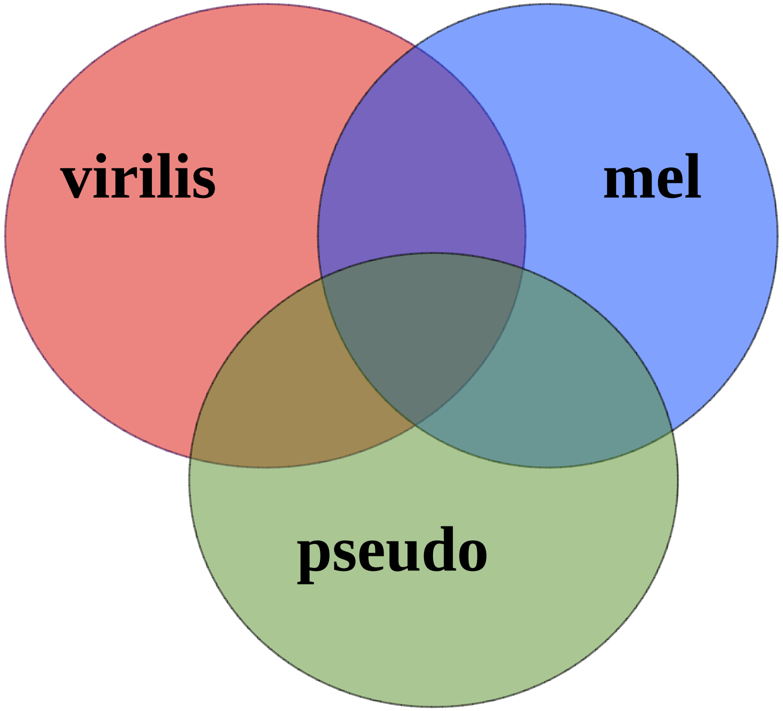
519 probes after cutoff for enrichment and p-value

Add 500 bp on either side of the probes

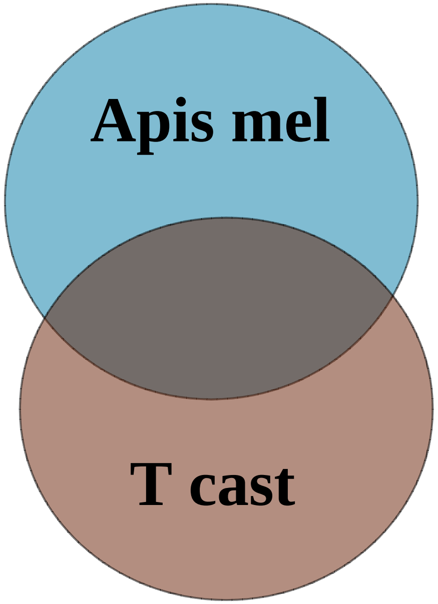
Extract aligning sequences from related species



Comparison of TFs in two-winged vs those in 4-winged



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, ANTP_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

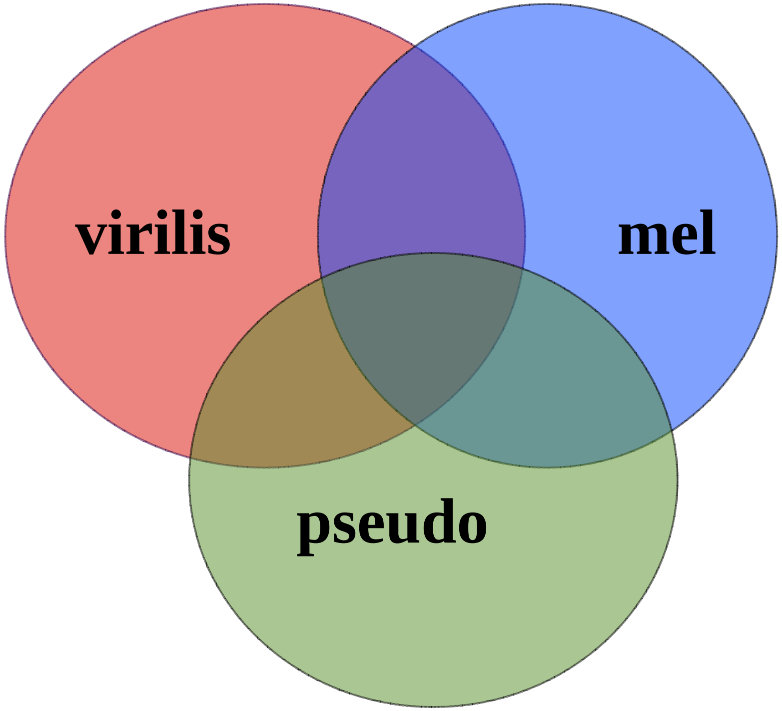
- Take region 2kb upstream of *pipsqueak* in *D melanogaster* and regions aligning to it, from *D pseudoobscura*, *D virilis*, *A mellifera*, and *T castaneum*
- Locate TFBSs in each sequence using the TRANSFAC Pro database (66 insect TF insect)

TFBSs (from TRANSFAC) between insects in the promoter region (1.5kb upstream) of *pipsqueak*

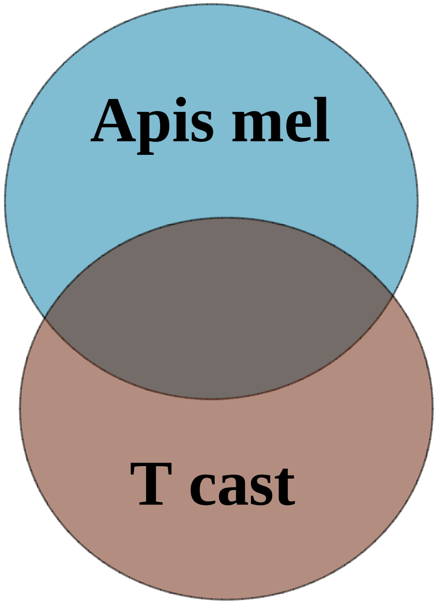
	Dmel	Dpse	Dvir	Amel	Tcas
Dmel	15(34)	8	11	6	6
Dpse	7	13(28)	8	7	5
Dvir	4	5	16(30)	7	7
Amel	9	6	9	16(34)	6
Tcas	9	8	9	10	8(17)

🟩 **Common** 🟧 **Difference (top-left)** ⬜ **TFs (total sites)**

Next slide compares all TFs in two-winged vs those in 4-winged

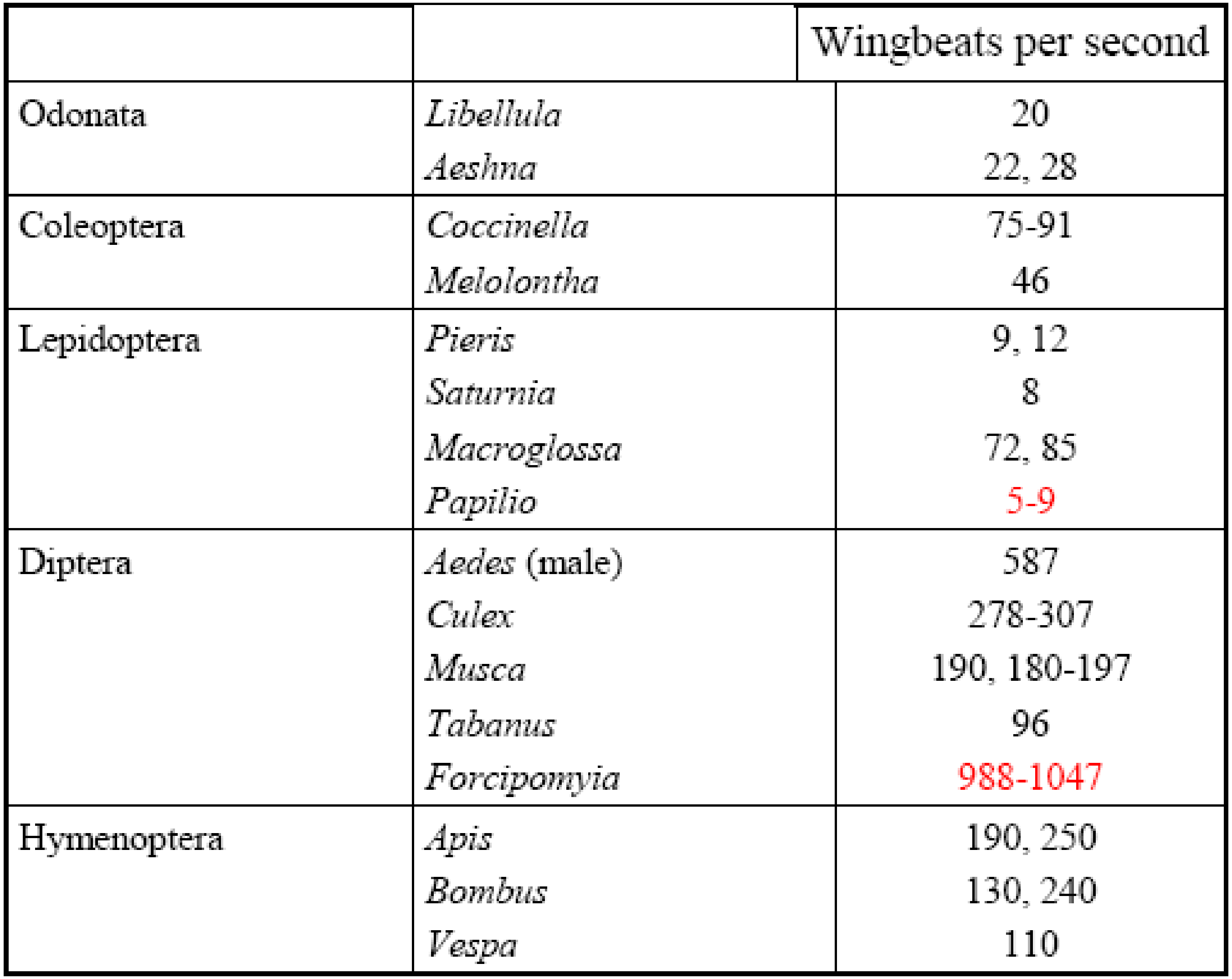


MAD_Q6, ADF1_Q6, ABDB_Q6, BRCZ3_01, HB_01, CAD_Q6, DEAF_01, SD_Q6, ABDB_01, BRK_Q6, DL_02, EVE_Q6, ZEN_Q6, DRI_01, FTZ_01, SGF3_Q6, UBX_01, SN_02, PRD_Q6, ABDA_Q6, BRCZ1_01, GAGAFACTOR_Q6, ANTP_Q6_01



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

What is the selection pressure for the evolution of two-winged insects?



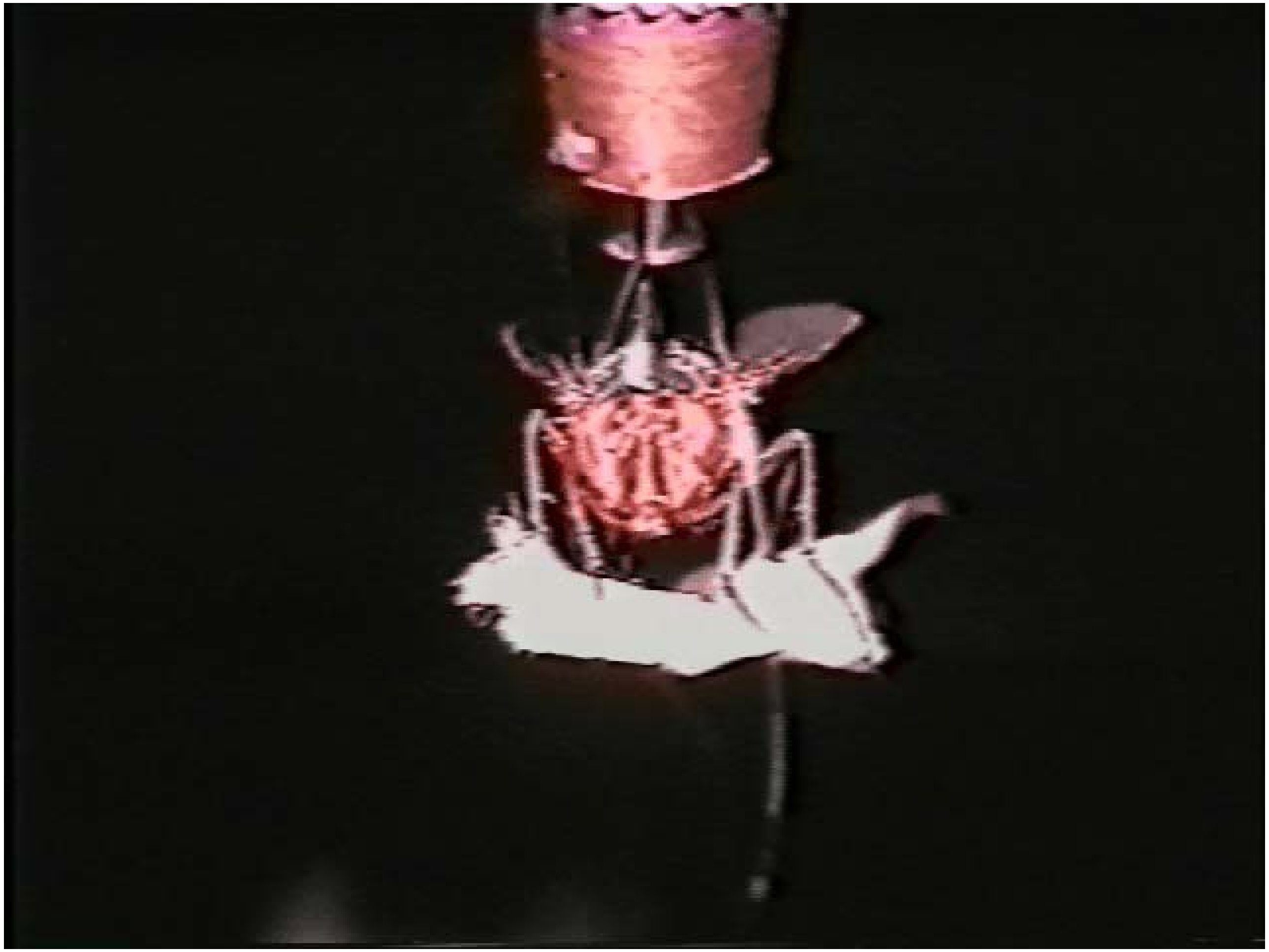
		Wingbeats per second
Odonata	*Libellula*	20
	Aeshna	22, 28
Coleoptera	*Coccinella*	75-91
	Melolontha	46
Lepidoptera	*Pieris*	9, 12
	Saturnia	8
	Macroglossa	72, 85
	Papilio	5-9
Diptera	*Aedes* (male)	587
	Culex	278-307
	Musca	190, 180-197
	Tabanus	96
	Forcipomyia	988-1047
Hymenoptera	*Apis*	190, 250
	Bombus	130, 240
	Vespa	110

Macroglossa

Forcipomyia

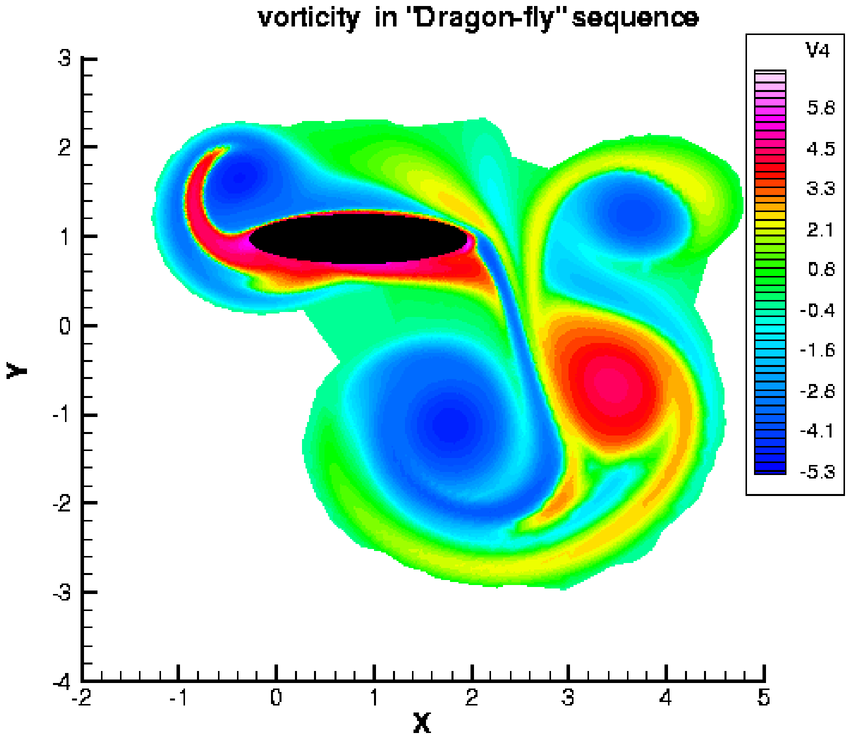
Tabanus

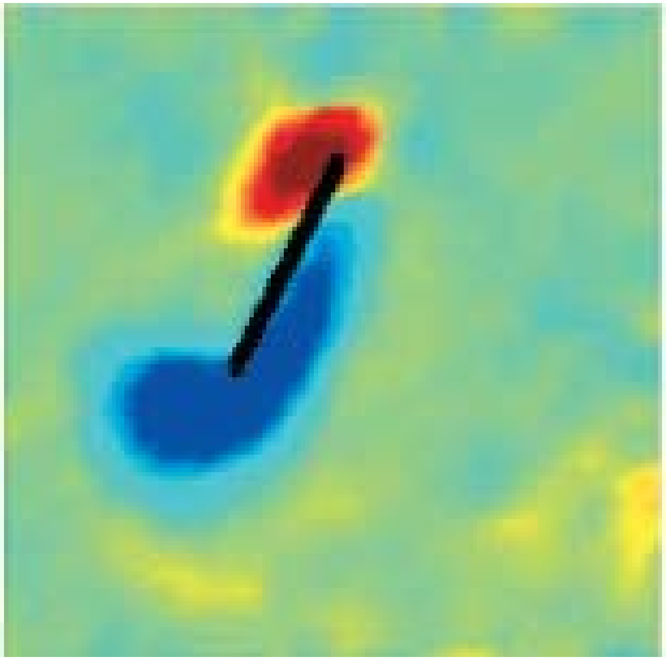
Drosophila beats its wings about 200 times per second.



CREDIT: Michael Dickinson

Vorticity in dragonfly sequence





Vorticity in a dipteran sequence