Hox genes and evolution of arthropod body plan: A comparative analysis of targets of Ultrabithorax in *Drosophila melanogaster* and *Apis mellifera*

A thesis submitted towards partial fulfilment of the requirements of the degree of Doctor of Philosophy by

Naveen Prasad

(20083006)

Indian Institute of Science Education and Research Pune

2013

To my Parents

CERTIFICATE

This is to certify that the work incorporated in the thesis entitled " Hox genes and the evolution of arthropod body plan: A comparative analysis of targets of Ultrabithorax in *Drosophila melanogaster* and *Apis mellifera*" submitted by Mr. Naveen Prasad was carried out by him under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other university or institution.

Date: Prof. L. S. Shashidhara

(Advisor)

DECLARATION

I declare that this written submission represents my ideas in my own words and where others ideas have been included; I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that the violation of the above will be the cause for disciplinary action by the institute and can also evoke penal action from the sources

Date: Naveen Prasad

(ID Number : 20083006)

Acknowledgements

I'd like to begin by thanking my advisor Prof. L.S. Shashidhara for his continued support through the course of my PhD training. I would also like to express my sincere gratitude towards the Director, CCMB for my initial stint at the institute which included the completion of my introductory course work. I'm grateful to IISER, Pune for providing me with the facilities required to conduct my research.

A major part of the work would not have been completed without the help of Prof. Martin Beye and Dr. Martin Hasselmann, Heinrich Heine University, Dusseldorf; both of whom with their lab members helped me with the biology of honey bees and provided me with enough honey bees to work with. I must also extend my sincere gratitude to Dr. Shashidhar Viraktmath, University of Agricultural Sciences, Dharwad for helping me with my preliminary studies on honey bees and for his help in acquiring local strains of honey bee.

Dr. Farhat Habib helped me greatly in downstream bioinformatics analysis of ChIP seq data. This analysis would have extended much longer if it wasn't for his valued advice and help.

I would also like to thank Dr. Anuradha Ratnaparkhi for her help with various fly techniques and for her valuable advice relating to fly genetics. She has always been very prompt with her help. I must mention the help and support extended by Dr. Girish Ratnaparakhi and Dr. Richa Rikhy to various aspects of fly biology and to the lab at IISER, Pune. The acknowledgement can never be complete without mentioning Dr. Surendra Ghaskadbi who helped me with his advice and questions.

I would also like to express my gratitude to the senior members of our lab , especially Ramakrishna and Sudha who helped me immensely during the initial phases of my PhD career. I would also like to thank many other members of the lab especially Arkaja, Savita, Julien, Harsha, Jay Prakash who have always made the lab a lively place. I would like to thank the Biology floor members at IISER

Pune (We started as a really small group ☺) for putting up with me for all of these years.

I would like to take this opportunity thank various funding agencies – CSIR,India; DAAD and The Company of Biologists who provided me with various grants at different times to advance my research work.

On a more personal note, I would like to thank Zenobia for her encouragement and support for all of these years- something that is needed most during PhD. I would also like to thank my brothers Nabish and Pawan who helped me immensely with regards to my career decisions. I would like to thank my parents for their much needed support in my personal life. It is to them that I attribute the person that I am today.

Synopsis

Title : " Hox genes and the evolution of arthropod body plan : A comparative analysis of targets of Ultrabithorax in *Drosophila melanogaster* and *Apis mellifera*"

Name of the student: Naveen Prasad

Name of theis advisor: Prof. L.S. Shashidhara

Indian Institute of Science Education and Research (IISER), Pune

Introduction

Amongst insects, there has been much diversity in the number and size of wings. Most insects have four wings (all directly contribute to the flight), while beetles and flies have only one pair of wings. In fruit flies, the hind wing is modified to a small bulbous balancing organ called haltere, while in the beetles the forewing is modified into a protective structure called elytron. It is now widely accepted that evolution at the level of a family of highly conserved (from insects to human) genes popularly known as Hox genes is one of the major factors for the diversity in animal body plan. Hox genes code for homeodomain containing proteins (reviewed by McGinnis *et. al.*,1992); which are expressed along the anterior posterior axis of the body and impart distinct identity to individual segments (Akam, 1987). Suppression of wing fate and specification of haltere fate in *Drosophila* by the Hox gene *Ultrabithorax* (*Ubx*) is one of the well-studied models for Hox regulation of development (Lewis, 1978). Earlier studies indicate that Ubx protein or its expression pattern has not evolved amongst the diverse insect groups, although they appear to regulate the differences between developing forewing and hindwing in at least three insect groups studied so far (Grenier *et. al.*, 2000; Galant *et. al.*, 2002; Ronshaugen *et. al.*, 2002).

Morphological differences between wing and haltere in *Drosophila melanogaster*

The wing is a flattened structure composed of veins and interveins. Veins are hollow structures through which the trachea and the nerve fibres run. The intervein region is formed by apposition of two layers and in the adult fly the cells in this region are dead. Haltere on the other hand is a bulbous structure and is much smaller than the wing. It is devoid of any vein intervein pattern and serves to detect the flow of air during the flight of fruit flies.

Morphological differences between fore and hind wing of *Apis mellifera*

The fore and hind wings of honeybee are almost identical except for a few differences. Unlike haltere of fruit fly, which is very different from the wing, the hind wing of honeybee is very similar to the forewing. There are subtle differences in the venation pattern between the fore and hind wing. An important difference that is seen in the fore and hind wing is the orientation of bristles on the edge of the wings. A set or interlocking bristles are found in the posterior end of the fore wing, while the complementary set is found in the anterior end of the hindwing. This interlocking mechanism serves to lock the fore and hind wings during flight.

The question therefore arises how Ubx is able to modify the fate of wing to various degrees in different insect orders. This work is part of a larger study aimed to understand variations in cis regulatory elements (CREs) of targets of Ubx to generate diversity amongst insects. The current work is aimed to understand developmental and molecular events downstream of Ubx resulting in differences in wing morphology between two divergent insect groups – *Apis mellifera* and *Drosophila melanogaster*. I used a genome wide approach to identify direct targets of Ubx in *Apis* and compared the same to those in *Drosophila*.

Objectives of the study

To understand the evolution of downstream events of Ubx resulting in differences in wing morphology; the direct binding sites of Ubx in honeybee need to be identified and subsequently compared to those in other insects to look for similarities and differences and finally reconstruct the genetic events that might have taken place while shaping a haltere instead of a hind wing in fruit flies. The objectives of the present study to achieve this end are summarised as:

1. Identification of direct binding regions of Ubx in the honeybee, *Apis mellifera.*
2. Identification of genes directly targeted by Ubx in hindwing disc of *Apis mellifera* and compare the same to those in halteres of *Drosophila melanogaster* and identify the diptera specific and the hymenoptera specific targets that might be involved in shaping the haltere or hindwing, respectively.
3. Analysis of the direct binding regions using bioinformatics tools in order to understand the mechanisms that have enabled Ubx to shape a haltere in fruit flies instead of a hindwing in honeybee.
4. Analysis of the genes that are Ubx targets in the two insects and understanding the probable role of diptera specific Ubx target genes in shaping the haltere.

Results & Discussion

Identification of direct binding regions of Ubx in the honeybee, *Apis mellifera*

Wing patterning network has remained conserved for over 300 million years in all the insects studied so far (Abouheif *et. al*., 2008). Expression of a few major wing patterning genes were studied in the wing discs of *Apis mellifera*. The expression of these genes, namely *Cut* (*ct*), *Spalt*(*sal*), *extradenticle* (*exd*)and *engrailed*(*en)*in the wing discs is identical to those in other hymenopterans. Also, there was no difference in expression pattern of these genes between the fore and hind wing

disc. The results suggest expression patterns of these genes are qualitatively identical between the fore and hind wing of honeybee (unlike in fruit flies, where there is a clear difference in the expression pattern of these genes between wing and haltere).

Based on expression of Ct in the wing disc of honey bee, the expression of which in the wing discs of fruit fly marks the late third instar larva of the fly; early fifth larval instar of honeybee was chosen for various chromatin immunoprecipitation (ChIP) studies.

In order to identify the direct binding regions of Ubx in the wing discs of honeybee, ChIP was performed on the wing discs of honeybee using anti-Ubx antibodies. For performing ChIP, polyclonal antibodies were raised against the N-terminus of Ubx protein (homeodomain and YPWM motif on the C- terminus end were excluded in the recombinant protein to avoid non specific interactions) by cloning and expressing the protein in bacteria and subsequently injecting the same in rabbit, followed by extraction of sera according to standard protocols. The specificity of the antibody was checked by Western blot analysis of whole embryonal lysate and immunohistochemistry on wing discs. On the Western blot, a single band corresponding to around 35 kDa was obtained which is the expected molecular weight of Ubx protein. Antibody staining on wing discs of honeybee showed higher levels of expression of Ubx protein in the hind wing discs as compared to the forewing disc. In both fore wing and hind wing discs, Ubx was predominantly localized to the nucleus.

ChIP was carried out on both fore and hind wing samples. In the honeybee hive, a single queen lays eggs and hence all the bees/ larvae present in a single hive are related to each other. For carrying out ChIP, artificially inseminated queen (using semen of a single drone) was used. Each replicate of ChIP was carried out using samples from one single hive (in order to minimise the heterogeneity amongst population). Three independent biological replicates of ChIP were carried out using samples from three different hives.

Identification of the direct target genes of Ubx in honeybee and comparison of the same to those in fruit fly

Sequencing of the chipped DNA was carried out on Illmina platform and the resulting sequences were checked for their quality using FastQC. Only the sequences that passed the quality testing were used for further analysis. The sequences were then aligned to the genome of *Apis mellifera* (version 2.0) and peaks (regions bound by Ubx) were identified using MACS. In order to ensure that only true peaks were reported, a stringent cut off for peak selection was set up. All the peaks were identified using input as control, following which all the pre IgG peaks were subtracted from anti-Ubx peaks. Peaks with a fold enrichment of 10 fold and above with a FDR value of <= 1% were selected for further analysis. Using these criteria, 2250 peaks (regions bound by Ubx) were identified in the hindwing discs of *Apis mellifera,* while 927 peaks were detected in the forewing discs of *Apis mellifera,* amongst which 526 were common to the two datasets.

A few of these peaks were verified using ChIP-qPCR. Nearest genes to each of these peaks were identified using bioinformatics. A total of 1396 target genes of Ubx were identified in the hind wing discs of honeybee, of which 1182 had known orthologs in *Drosophila*. 583 genes were identified as direct targets of Ubx in the forewing discs of honeybee, of which 528 had known orthologs in the fruit fly. The hindwing targets of Ubx were compared to the Ubx targets in fruit fly from two different studies (Agrawal *et. al*., 2011 & Choo *et. al*., 2011).

On comparison of either dataset with honey bee Ubx targets, three types of gene sets emerged- one is diptera specific gene set, second is gene set that is common to both diptera and hymenoptera and finally a hymenoptera specific gene set. The genes of these sets were assigned to various biological processes and signalling pathways and few important ones were picked up for comparison. It was found that all the major biological processes and signalling pathways that are targeted by Ubx in dipterans are also targets of Ubx in honey bees. However, few biological processes like wing disc development, cell adhesion, growth and its regulation are associated with a much higher proportion of diptera specific target

genes than the hymenoptera specific targets. It may be hypothesized that these genes that are under the control of Ubx in the dipteran lineage have played a very important role in shaping a haltere instead of a hindwing.

In order to test relative importance of (a) genes that are conserved as targets of Ubx over the long evolutionary history, and (b) genes have come under the influence of Ubx only in dipterans, a comparison was made between microarray data and the direct targets of Ubx genes. It was found that both the diptera-specific target genes and the common target genes are represented at roughly equal proportion in the microarray data. This suggests that both the types of Ubx targets- ancestral Ubx targets and the novel target genes of Ubx in halteres have played an equally important role in shaping the haltere instead of wing. One therefore has to understand the role of diptera specific Ubx target genes in shaping a haltere and the way in which the pre-existing targets of Ubx have together function to shape a haltere instead of a hind wing.

Antibody staining of few important wing patterning genes, which are Ubx targets in both the hind wing of honeybee and halteres of fruit fly showed identical expression patterns in the fore and hind wing discs of honeybee, while the same genes are down regulated in the halteres when compared to wing discs of *Drosophila*. It is possible that pre-existing targets are regulated by Ubx in a novel way to shape the haltere.

Analysis of the ChIPped sequences to identify mechanism of gene regulation by Ubx in honeybee and comparison of the same to those in fruit fly

To understand what changes might have taken place in the regulatory regions bound by Ubx between the two species, motifs were identified in the ChIPped sequences of honey bee and were compared to those already identified in Ubx bound regions in the fruit fly. We found that Ubx core binding motif (TAAT) was under represented in the pulled down sequences when compared to the random sequences (similar results have been reported for fruit fly). Furthermore, patterns of distribution of binding sites for other transcription factors/chromatin modulators motifs are similar between honeybees and fruit flies. It is, therefore,

possible that mechanism by which Ubx recognizes its downstream targets is similar between the two insects.

Closer examination of the Ubx-binding sites of such genes suggested that Ubx modulates the activity of its target genes by associating with various co-factors in a context specific manner, which are present in the vicinity of Ubx-binding regions in diptera and such motifs are absent in honeybees. We, therefore, suggest that while large number of targets of Ubx are similar between honeybee and *Drosophila*, in the latter that regulation of transcription is more pronounced due to the acquisition of binding sites for co-factors of Ubx. This hypothesis is being tested now.

Characterization of a diptera specific Ubx target gene- *Gliolectin*

Gliolectin (Glec) is a target of Ubx in Drosophila, but not in honeybee. In order to understand the role played by diptera-specific Ubx targets in haltere development, functional characterization Glec was undertaken. *Glec* is expressed at the D-V boundary of wing imaginal discs and also along the prospective interveins. The use of protein trap lines showed that its expression is maximal along the D-V boundary and the interveins and is less in the areas where Delta is expressed. Its expression is highly reduced in haltere discs. Genetic experiments using various UAS- *Glec* and UAS $Glec^{RNAi}$ lines showed that:

Loss of function phenotypes of *Glec* mimic loss of Notch function (or gain of EGFR activity).

Gain of function phenotypes of *Glec* mimic gain of Notch (or loss of EGFR activity) function.

Loss of function of Glec also downregulated target genes of Wg such as *Delta* and *Ach* in the wing imaginal disc, while gain of function of Glec induced the expression of Wg at the DV boundary of haltere in posterior compartment without affecting Delta expression.

Above results, thus, suggest that Glec may modulate Notch activity. Possible localisation of Glec in the golgi suggests that the protein is necessary for Notch in

order to respond to its ligands- Delta and Serrate, which is currently being explored.

References:

Abouheif, Ehab, and Gregory A Wray. Evolution of the Gene Network Underlying Wing Polyphenism in Ants. *Science* 297 (2008): 249–252.

Agrawal P, Habib F., Yelagandula R and Shashidhara LS. Genome level identification of targets of hox protein Ultrabithorax in Drosophila : Novel methods of target selection. *Scientific Reports* 205 (2011):

Averof, M, and N H Patel. Crustacean Appendage Evolution Associated with Changes in Hox Gene Expression. *Nature* 388.6643 (1997): 682–6.

Carroll, Sean B, Scott D Weatherbee, and James A Langeland. Homeotic Genes and the Regulation and Evolution of Insect Wing Number. *Nature* 375 (1995): 58–61.

Castelli-gair, James, and Michael Akam. How the Hox Gene Ultrabithorax Specifies Two Different Segments: The Significance of Spatial and Temporal Regulation Within Metameres. *Development* 121 (1995): 2973–2982.

Choo SW, White R and Russell S . Genome wide analysis of the binding of the Hox protein Ultrabithorax and the Hox co factor Homothorax in *Drosophila*. *Plos One* 6(4). 2011

De Celis, J F, S Bray, and a Garcia-Bellido. Notch Signalling Regulates Veinlet Expression and Establishes Boundaries Between Veins and Interveins in the Drosophila Wing. *Development (Cambridge, England)* 124.10 (1997): 1919–28.

Grenier *et. al.* Functional evolution of the Ultrabithorax protein. *PNAS.* 97(2). 2000: 704- 9

Galant *et. al.* Evolution of transcriptional repressor domain in an insect Hox protein. *Nature* 415 (6874) 2002: 910-3

Huang, Da Wei, Brad T Sherman, and Richard A Lempicki. Bioinformatics Enrichment Tools: Paths Toward the Comprehensive Functional Analysis of Large Gene Lists. *Nucleic acids research* 37.1 (2009): 1–13.

Hueber, Stefanie D *et. al*. Comparative Analysis of Hox Downstream Genes in Drosophila. *Development* 134 (2007): 381–392.

Mahfooz, Najmus *et. al.* Ubx Regulates Differential Enlargement and Diversification of Insect Hind Legs. *PloS one* 2.9 (2007): e866.

Makhijani, Kalpana *et. al.* Modulation of Decapentaplegic Gradient During Haltere Specification in Drosophila. *Developmental biology* 302.1 (2007): 243–55.

McGinnis, W *et. al.* A Homologous Protein-coding Sequence in Drosophila Homeotic Genes and Its Conservation in Other Metazoans. *Cell* 37.2 (1984): 403–8.

Mohit, Prasad *et. al.* Modulation of AP and DV Signaling Pathways by the Homeotic Gene Ultrabithorax During Haltere Development in Drosophila. *Developmental biology* 291.2 (2006): 356–67.

Roshaugen *et.al.* Hox protein mutation and macroevolution of the insect body plan. *Nature.* 415 (2002): 209-15

Slattery, Matthew *et. al.*. Cofactor Binding Evokes Latent Differences in DNA Binding Specificity Between Hox Proteins. *Cell* 147.6 (2011): 1270–1282.

Slattery M, Ma L, Negre N, White KP and Mann R. Genome-wide tissue- specific occupancy of the Hox protein Ultrabithorax and the Hox cofactor Homothorax in *Drosophila. Plos One* 6(4) (2011).

Weatherbee, S D *et. al.* Ultrabithorax Function in Butterfly Wings and the Evolution of Insect Wing Patterns. *Current biology : CB* 9.3 (1999): 109–15.

Weatherbee, Scott D *et. al.* Ultrabithorax Regulates Genes at Several Levels of the Wing-patterning Hierarchy to Shape the Development of the Drosophila Haltere. *Genes & Development* 12 (1998): 1474–1482.

Zhang, Yong *et. al*. Model-based Analysis of ChIP-Seq (MACS). *Genome biology* 9.9 (2008): R137.

Table of contents

Chapter 4

Comparison of direct target genes of Ubx in hind wing discs of *Apis mellifera* with those in halteres of *Drosophila melanogaster*

Chapter 5

Comparison of motifs found in Ubx bound regions of Apis mellifera to those in Drosophila melanogaster

Chapter 6

Characterization of *Gliolectin*, a diptera specific target of Ubx

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of tables

List of abbreviations

1. Anterobithorax	abx
2. Contrabithorax	Cbx
3. Bithorax	bx
4. Bithoraxoid	bxd
5. Postbithorax	pbx
6. Ultrabithorax	Ubx
7. Homeobox	Hox
8. Extradenticle	Exd
9. Ultrathorax	Utx
10. Bicoid	Bcd
11. Hunchback	Hb
12. Kruppel	Kr
13. Wingless	wg
14. Engrailed	En
15. Hedgehog	Hh
16. Epidermal growth factor receptor	EGFR
17. Short gastrulation	Sog
18. Rhomboid	Rho
19. Decapentpelagic	Dpp
20. Patched	ptc
21. Knot	kn
22. Optiomoter blind	omb
23. Vestigial	vg
24. Spalt	sal
25. Apterous	Ap
26. Vein	Vn
27. Fringe	Fng
28. Serrate	Ser
29. Delta	Dl
30. Cut	Ct

31. Distalless	Dll
32. Achaete	Ach
33. Chromatin Immunoprecipitation	ChIP
34. Phosphate Buffered Saline	PBS
35. Phosphate Buffered Saline (with Triton X-100)	PBTX
36. Phosphate Buffered Saline (with Tween- 20)	PBST
37. Bovine serum albumin	BSA
38. Tris Buffered Saline	TBS
39. Anterior- Posterior	A-P
40. Dorsal- ventral	D-V
41. Cis-regulatory elements	CRE
42. Tris acetate EDTA	TAE

Chapter1

Introduction

While polarity at the cellular (for example, apico-basal polarity) and organismal level (for example, oral-aboral) evolved much earlier, Urbilaterians (the common protosome- deuterosome ancestor or the PDA) are considered to be the first organisms that show a segmented body plan. There is a good amount of debate about the complexity of these Urbilaterians, with some authors suggesting a very simple body plan for these PDAs (based on various morphological and paleontogical evidence) (Erwin *et. al*., 2002) while others are of the view that these common ancestors already had complex body plan with a distinct D-V axis (Balavoine *et. al.*, 2003). Amongst the organisms that originated from these Urbilaterians, arthropods are the first group that show remarkable diversity in their body plan, both along the antero- posterior and dorso- ventral axis.

1.1: Genetic changes underlying the diversity in morphology of arthropods

The phylum arthropoda is considered to be one of the most successful phylum that has evolved to date. It consists of over one million named species (Degaard, 2000) and many feel that many more species are yet to be identified. These arthropods are unique in the sense that they inhabit all the ecosystems. The members of this phylum are sub classified into two subphyla namely- Mandibulata and Chelicerata. The subphylum Mandibulata is further sub classified into Myriapoda and Pancrustacea (which includes the group Hexapoda- to which insects belong) (Regier *et. al*., 2010). Overall, this phylum shows maximum diversity in terms of morphology and the habitat they inhabit.

The evolution of diversity in the morphology of arthropods, which must have played an important role in their adaptation to virtually every type of known habitat, has always aroused the interest of biologists- evolutionary biologists in particular. An important question that arises is how the diversity in the morphology has evolved. Differences in morphology have been attributed to the changes in genetic network of organisms. The question about understanding the diversity in morphology therefore

comes down to understanding the genetic basis or the molecular mechanisms underlying the morphological changes.

The morphology (together with other attributes of individual organisms) changes with time and most of these changes are adaptive in nature and hence they get fixed in the population. However, it seems that most of the genetic changes that take place confer no selective advantage to the individual carrying them (Kimura, 1983) - a theory popularly known as the theory of neutral selection. Thus, a very obvious and challenging task is reconstructing the sequence of events where the molecular changes might have led to adaptive morphological differences.

1.2: Homeotic genes

1.2.1: Hox genes as regulators of body plan along the anteroposterior (A-P) axis.

Homeosis, the term coined by Bateson, means "similar condition" and is associated with formation of a body part having features that are normally found in a related or similar part at another location in the body (reviewed in McGinnis, 1994). Genes associated with such homeotic mutations are called homeotic genes. In *Contrabithorax* (*Cbx)* mutants (White *et. al.*, 1985), for example, the wing of *Drosophila* is transformed into a haltere. The Hox genes serve to modify the basic or ground genetic framework of organism's segment and impart distinct identity to various segments (Akam, 1987). Since the formation of a single structure like the wing requires thousands of genes acting in a concerted manner, it is assumed that the homeotic genes such as *Ultrabithorax* (*Ubx*) are considered as master control genes. All homeotic genes encode for transcription factors with a conserved DNA-binding domain called homeodomain or homeobox. Most accepted name for this class of genes/proteins is Homeobox genes or in short Hox genes. All Hox genes are important developmental regulators, acting together to determine the identity of segments along the anterior- posterior axis.

Plate 1.1: Hox genes in *Drosophila melanogaster*

In the absence of all Hox genes, all segments develop the "same ground pattern," a mix of thoracic and cephalic pattern with no morphological diversity along the AP axis (Struhl, 1982).

1.2.2: Clustering of Hox genes on chromosomes and posterior prevalence

Hox genes are mostly clustered on chromosomes, though different types of Hox clusters are found in the animal kingdom. In *Drosophila*, for example, eight Hox genes are arranged in two clusters; both are located on 3rd chromosome. The genes provide identity to various segments along the anterior- posterior axis of the organism in the same order in which they lie on the chromosome (Sanchez Herrero *et. al*., 1985; Kaufman *et. al.*, 1990). The mechanistic basis of this co-linearity still remains a mystery. Hox genes also interact with each other to maintain a proper expression domain by both positive and negative regulation. A term called posterior prevalence is used to describe a phenomenon by means of which a more posterior gene suppresses the expression and function of the anterior genes. Initial observations that were made showed that genes of Bithorax complex were capable of down regulating *Antennapedia (Antp)* (Hafen et. al., 1994). Struhl et. al. in 1985 also showed that *Ubx* could be repressed by the more posteriorly expressed genes like *Abd-A* and *Abd-B*.

In flies, Hox genes are clustered into antennapedia and bithorax complexes. Some of the Hox genes in these clusters viz., *Zen, Zen2, Bcd, Ftz, Eve* have adapted novel functions by changing their structure and function and are no longer functional Hox proteins even though they are homeodomain containing proteins (reviewed in McGinnis *et. al*., 1992). In insect genomes, other than *Drosophila* such as *Schistocerca, Anopheles, Bombyx, Tribolium* and *Apis*, a single Hox gene complex exists and Hox genes within this cluster are arranged in the same collinear manner as in flies (Stuart *et. al*.,1991; Ueno *et. al*.,1992; Ferrier *et. al*.,1996; Power *et.al*.,2000). Mammalian genomes encode four Hox clusters while many teleost fishes have seven partial Hox clusters. Hox genes in mouse (Duboule *et.al*., 1989) and human (Acampora *et.al*. 1989) too show colinearity. Furthermore, the orthologs of *Drosophila* Hox genes are arranged in the same order in mice and human genome as they are found in fly genome.

Detailed sequence analysis suggest that vertebrate Hox clusters have arisen from an ancestral cluster containing just three Hox genes, which have diversified in due course of evolution by duplication and divergence.

Plate 1.2: Conserved Hox gene organisation and expression across various organisms.

1.2.3: Structure of Hox proteins

All the Hox proteins contain a 60 amino acid homeodomain, a DNA binding motif (reviewed by Scott *et. al*., 1989). Upstream of this homeodomain is a hexapeptide motif called the YPWM motif, which is known to interact with Extradenticle (Exd). Homeodomains when swapped between two Hox proteins, show similar specificity of DNA recognition (Johnson *et. al*., 1995; Chan *et.al.,* 1996). In general, there are three helices are in the homeodomain. Helix 3 binds to the major groove and makes majority of the contact between the DNA and protein. The remaining contacts are made by the N- terminal arm of the homeodomain. All the homeotic proteins, even though they have similar DNA binding domains exert unique functions in vivo. From various recent studies, it has become clear that these Hox proteins use various co-factors to exert specificity of binding to target DNA (Taghli- Lamallem *et. al*., 2007).

1.2.4: Regulation of Hox genes

Hox genes, which control a variety of downstream targets are themselves regulated in a complex manner. Various different levels of Hox gene regulation have been postulated which include- nuclear compartmentalization, chromatin elements, enhancer and promoter interactions, post-transcriptional and post-translational modifications.

In *Drosophila*, the Hox genes are first transcribed in the blastoderm, primarily under the regulation of gap and pair-rule genes (Dessain *et.al.,* 1991). Their early transcription patterns evolve rapidly, reflecting the dynamic generation of pattern in the blastoderm, but by the end of gastrulation, each of the homeotic genes is expressed in a defined spatial domain encompassing one or more segment primordia (Harding *et. al*., 1985; Carroll *et. al*., 1986; Mahaffey *et. al*., 1987; LeMotte *et. al*., 1989; DeLorenzi *et. al*., 1990; Karch *et. al*., 1990). These expression domains are maintained throughout subsequent development in the absence of the segmentation

gene products; by a memory mechanism that involves genes of the Polycomb and Trithorax groups (Paro, 1990; Kennison, 1993).The Polycomb and Trithorax elements help in maintaining the expression state of the Hox genes (Mihaly *et. al.*, 1998 and Ringrose *et. al.*, 2004). In general, the repression of Hox gene expression is maintained by the Polycomb family of proteins while their activation is regulated by the Trithorax group of proteins (Ingham, 1998; McKeon *et. al*., 1994; Simon, 1992). The chromatin boundary elements also demarcate the expression of Hox genes by preventing the cross talk between neighbouring regulatory elements. The micro RNAs and co factors are involved in fine-tuning the expression of Hox genes.

1.3: Regulation of arthropod body pattern by Hox genes

Hox genes regulate arthropod body patterns (and the body patterns of other phyla as well) and it may seem logical to assume that the evolution of arthropod body pattern is linked to the evolution of Hox genes. Several models have been suggested which link the evolution of body plan with the evolution of Hox genes. These include:

1. Changes in number of Hox genes (by duplication & divergence).
2. Changes in domain of Hox gene expression.
3. Changes in Hox gene that gives the protein its new properties.
4. Changes in Hox- protein responsive elements of downstream genes.

Change in Hox gene number has played an important role in shaping and evolution of body plan of different organisms. Urbilaterians, the ancestors of bilaterians are believed to have possessed a collinear cluster of atleast eight Hox genes (Rosa *et. al*., 1999 and Garcia –Fernadez *et. el.,* 2005). Insects, belonging to Protosomes too have a cluster of around eight hox genes. The evolution of vertebrates (belonging to deuterosomes) was preceded by the evolution of four Hox complexes (Holland *et. al*., 1998). It is believed that the change in Hox gene number, imparted the huge amount of much needed complexity in the chordate lineage.

Plate 1.3: Evolution of Hox gene complex in metazoans

It is possible that a number of Hox genes that arose by gene duplications might have very similar or in cases largely overlapping functions in the ancestral organisms. In the course of evolutionary history, these genes acquired new functions. One can

therefore assume that segmental diversity may not be associated with acquisition of new genes (by gene duplications), but by diversification of function of each gene.

Centipedes and onychoporans have a very different and relatively simpler body plan as compared to the more recently evolved arthropods. The trunk of these organisms consists of series of identical segments. However, studies have shown that these organisms also have eight Hox genes identical to those in the fruit fly, *Drosophila melanogaster*. *Artemia* (brine shrimp) has a very different pattern of segmental specialisation; still it carries the same set of Hox genes (Averof *et. al.*, 1993). Very similar types of Hox genes are also present in horse shoe crab (chelicerates) (Catwright *et.al.*,1993). Thus, complexity in the body plan of these organisms does not appear to be associated with the divergence of Hox genes.

Diversification in function of Hox genes can be achieved by either diversification in the expression pattern of various Hox genes or by changes in regulation of downstream targets by Hox genes in individual organisms. It seems that a correlated shift in the pattern of Hox gene expression caused some organisms (maxillipeds) to have two pairs of specialised front legs while the brine shrimps lost them (Averof *et. al*., 1997). Differences in expression pattern of Hox genes may also explain why the insect body is divided into head, thorax and abdomen while the centipedes have a long homologous trunk (Grenier *et. al*., 1997 and Hughes *et. al.*, 2002).

Plate 1.4: Segmental specialization and *Ubx* expression amongst various chelicerates

The expression of Hox genes has remained fairly identical in various insect orders but still there is considerable amount of morphological differences which cannot be accounted for by either change in expression pattern of Hox genes or by change in number / diversification of Hox genes themselves. The fact that all the insects share identical Hox genes, whose expression pattern remains the same across insect orders but still a large amount of diversity exists amongst members of this order suggests that evolution at the levels of the downstream targets of Hox genes may have brought about such diversity amongst insects. Hox genes, being transcription factors, may regulate novel genes/ other factors such as micro-RNAs, if the latter acquire Hox-

binding sites in their enhancers. Hox genes may also take the help of new cofactors to bring new targets under their control or to modify the existing targets.

The current study deals with only insects as the expression pattern of various Hox genes and the number of Hox genes is same in all the members of this order. In addition, various Hox proteins appear to be functionally conserved amongst members of the order. Therefore, the current study limits itself to the study of changes / modification in binding sites of hox genes and their role in modifying the body plan.

The cis – regulatory elements (CREs as they are called in short) are the elements that are bound by various transcription factors and regulate the activity of one or multiple genes. These CREs consist of binding sites for various transcription factors. Binding sites (also termed as motifs) for individual transcription factor are simple and short (typically 4- 10 nucleotides long). The motifs often show degeneracy in the sequence and in a given CRE, the arrangement of motifs may vary from one species to a closely related species. Thus, CREs change more rapidly during evolution compared to coding sequences. It likely that the modularity and dispersal of CREs may remain neutral in the short term but can be important for evolution in the long term.

In the current study on the evolution of insect body plan, we examined CREs responding to the presence or absence of a Hox protein.

An Evo- devo approach to the understanding the role of Hox genes in shaping a novel organ using the case study of differential development of dorsal second and third thoracic organ between two insects (honey bee and fruit fly)

Suppression of wing fate and specification of haltere fate in *Drosophila* by the Hox gene *Ubx* is a classical example of Hox regulation of serial homology, which has served as a paradigm for understanding Hox gene function (Lewis, 1978). Interestingly, Ubx protein itself has not evolved amongst the diverse insect groups, although there are significant differences in Ubx sequences between *Drosophila* and crustacean arthropods (Galant *et. al*., 2002; Ronshaugen *et. al.,* 2002). Nevertheless, over-expression of Ubx derived from other insect group such as Lepidoptera is

sufficient to induce wing-to-haltere transformations in *Drosophila* (Grenier *et. al.*, 2000).

During the course of evolution; the expression and function of Ubx has remain conserved in various insects (as outlined later). However, the third thoracic segment which is shaped by Ubx has been modified to various degrees in various insects. Though the dorsal second thoracic organ is generally a wing in most insects, the dorsal third thoracic organ morphology varies a lot when compared to its second thoracic counterparts- a haltere in fruit fly while almost a similar hind wing in dragon flies.

The question therefore arises how Ubx is able to modify the fate of wing (second thoracic organ) to various degrees in different insect orders. It has been suggested that in the dipteran lineage, certain wing patterning genes have come under the regulation of Ubx. It has also been postulated that Ubx regulates its target genes or brings novel target genes under its control by binding to their enhancers (either by itself or with the help of other cofactors).

This work aims to identify the direct targets of Ubx in hind wing of honey bee and compare the same to those in fruit flies, therefore helping us to identify the fruit fly specific Ubx targets which might have helped shape the haltere in the dipteran lineage. This work also aims to understand variations in CREs of targets of Ubx to generate diversity amongst insects. Essentially, the study aims to identify the direct binding sites (CREs) of Ubx in few insects and compare the same to those in fruit fly. Analysis of the motifs present in these regulatory sequences will give us an idea about how various mutations have accumulated and ultimately changed the morphology of the wing to haltere. Further, by identifying genes associated with these CREs (using in silico approach); one can also understand which genes have been modified in dipteran lineage specifically and how these genes are important for shaping the haltere. Here, we report identification of targets of Ubx in *Apis mellifera*, the European honey bee, and their comparison with targets of Ubx in *Drosophila*. (**plate1.5**)

1.4: Role of Ubx in imparting diversity to the third thoracic segment of insects

A common feature that identifies the insects is the presence of three pairs of thoracic legs and two pairs of thoracic wings. The wings are found on the dorsal side of the body while the legs on the ventral side. The wings on third thoracic segment (which itself is a modification of the second thoracic wing) are modified to different extents in different insects.

In the ancient insects like dragon flies (belonging to the order odonata), the fore and hind wing are almost similar in shape and size; even though *Ubx* is expressed in the third thoracic segment.

In the common silk moth, *Bombyx mori*, the fore and hind wings are again similar; *Ubx* being expressed only in the hind wing. In honey bee, which is the model organism of this work, the posterior wing pair differs from the anterior in being slightly smaller and having a different venation pattern.

In butterflies, the fore and hind wings differ in their ornamental pattern and this change is brought about by the action of Ubx. It has been demonstrated that loss of *Ubx* in clones of hind wing results in transformation of the clonal regions to their fore wing counterparts (Weatherbee *et. al.* , 1999)

Interestingly, identity of second thoracic dorsal appendage in beetles, namely the elytron is changed to functional wing in the third thoracic segment by Ubx. RNAi analysis has shown that *Ubx* is essential for the formation of wings in the third thoracic segment instead of an elytron in the second thoracic segment of beetles. (Tomoyasu *et. al.*,2006)

Few studies have also tried to uncover the role of Ubx in the development of T3 legs. In the milkweed bug (*Oncopeltus fasciatus*) and house cricket (*Acheta domesticus*) the third thoracic legs are moderately and greatly enlarged in size respectively. By using RNAi technique to down regulate *Ubx* in these insects, it has been shown that Ubx is essentially responsible for diversification of the legs in the T3 segment while suppressing the formation of legs in the first abdominal segment (Mahfooz *et. al*., 2007).

Striking dimorphism is observed in the third thoracic legs of worker and queen honey bee, both of them being diploid females. While the third thoracic legs of the worker bees has an expanded bristle free region for carrying pollen etc called the corbicula, in the queens this structure is absent. The dimorphism is attributed to the difference in the expression levels of Ubx (worker hind legs expressing around 25 fold more Ubx than queen hind legs) and its localisation (Bomtorin *et.al.*, 2012).

In *Drosophila*, Ubx is both sufficient and necessary for transformation of wing to haltere. Ectopic expression of *Ubx* in the second thoracic segment is sufficient to cause transformation of wing to a haltere while the loss of *Ubx* in the third thoracic segment leads to the development of wing instead of haltere (Lewis, 1978; Cabrera *et. al.*, 1985; White and Akam, 1985). The protein Ubx has been shown to be needed at various stages of the development of *Drosophila* for formation of a proper haltere (Roch *et. al*., 2002).

Plate 1.6: Modifications of wing (dorsal second thoracic structure) to various degrees in different insects.

1.5: Expression of Ubx in various insects

Ubx is expressed in *Drosophila* embryos from posterior thorax through the abdomen. It is expressed mostly in the posterior part of T2 and T3 in thorax and in abdomen its expression extends to A8. In the posterior of each abdominal segment, the expression of *Ubx* is lowest (Akam & Arias, 1985; Akam *et. al.*, 1985; Carroll *et. al.*, 1988; Arias *et. al*., 1988).

Very similar pattern of *Ubx* expression is seen also in *Apis mellifera* embryos (Walldorf, 2000). *Ubx* is expressed maximally in the A1 segment during germ band stage of honey bee embryo whereas it can also be detected in the abdominal segments of the intermediate germ band stage embryos. The expression of *Ubx* is maximal around the tracheal pits, weak along the segmental boundaries and absent in the ventral part of the embryos.

Ultrathorax (ortholog of *Ubx* in *Tribolium*) expression begins in the early germ band stage where it is detected in the anterior border of fourth parasegment. In the mid

germ band stage, Utx expression decreases in parasegment four and in the later stages, it become undetectable in the parasegment. At the completion of germ band elongation stage, the Ultrathorax transcripts are detected from parasegment five through parasegement sixteen. The highest levels of staining are detected in the first abdominal segment (Bennet *et. al.*, 1999).

Studies carried out on embryos of *Precis coenia* (butterflies) to understand the expression of *Ubx* show maximal expression of *Ubx* in the first abdominal segment. Lower levels of Ubx are also detected in the second and third thoracic segment (Warren *et. al.*, 1994).

In the silk worm, *Bombyx mori, Ubx* expression has been detected from third thoracic segment to the ninth abdominal segment. However, in the later stages, the expression could also be detected in the second thoracic segment and extended through the ninth abdominal segment. Maximal expression of *Ubx* was seen in the first abdominal segment (Masumoto *et. al.*, 2009).

(Expression pattern of Ubx in various insect embryos have been summarised in **plate 1.7**)

The expression of *Ubx* in the wing imaginal discs of *Drosophila* is limited to the peripodial membrane while it is detected almost everywhere in the halteres (Akam, 1983; White and Wilcox, 1984). The current study describes the presence of Ubx in both fore and hind wing buds of *Apis mellifera* but quantitatively more Ubx is detected in the hind wing bud when compared to the fore wing bud. Expression of *Ubx* in *Tribolium castaneum* in the wing is much more as compared to that in the elytron. (Tomoyasu *et. al*., 2005). *Ubx* expression has also been studied in the wing buds of butterflies (*Precis coenia*) and Ubx is completely absent in the wing discs (Warren *et. al.*1994) but reports do suggest that Ubx is present in the peripodial membrane of the forewing.

(Expression pattern of Ubx in various insect orders has been summarised in **plate 1.8**)

1.6: Functional evolution of Ubx

Various functions of the protein Ubx have remained conserved in the long evolutionary history. In order to understand the functional similarity between Ubx orthologs, a comparison of Onychoporan Ubx and *Drosophila* Ubx was carried out. Onychophora is a sister phylum of arthropoda and the protein Ubx has been diverging in sequence for about more that 520 million years. Ectopic expression of onychophoran Ubx carried out almost the same functions in *Drosophila melanogaster* as its own Ubx. However, some of the embryonic functions of Ubx in *Drosophila* couldn't be mimicked by the Onychophora Ubx (Grenier *et. al.,* 2000). However, most of the embryonic functions of Ubx in *Drosophila* were mimicked by other insect Ubx like *Tribolium and Junonia*.

To understand the evolution of Ubx protein in the insect lineage when compared to the oncyphoran Ubx, a study was carried out which compared the Ubx proteins of various insects and an onychophoran, *Akanthokara kaputensis*. Ubx proteins in the insect lineage have acquired a QA domain towards their C terminal end over the course of evolution. This domain present in all the insects plays an important role in carrying out some of the repressive functions of Ubx in the embryonic stage (Galant *et. al.*, 2002).

1.7: General structure and organisation of *Ubx* locus

Bithorax complex in *Drosophila melanogaster* contains three genes of which *Ubx* is the anterior most expressed gene. *Ubx* along with other two genes of the complex – namely *Abd A* and *Abd B* specify nine distinct segment types. It is therefore evident that each of these Hox genes can essentially modify more than one segment pattern. All the genes of bithorax complex are expressed within parasegments, a parasegment consisting of the posterior compartment of one segment and anterior part of the other. The genes of the complex are responsible for imparting identity to parasegment 5-parasegment 14 out of the 14 parasegments present in fruit fly. The function of *Ubx* is best understood amongst all the genes of this complex.

Ubx domain plays primary role in imparting identity to PS5 and PS6 (especially in embryo and larval cuticle), it also has some functions in the epidermis of more posterior segments. Pit sense organs in the abdominal segments are suppressed by Ubx and normal abdominal setae develop because of Ubx (Sthrul, 1984 & Lewis, 1978). Four different mutations have been described for *Ubx*, namely- anterobithorax (abx), bithorax (bx), bithoraxoid (bxd) and postbithorax (pbx). The abx and bx mutations transform the PS5 adult cuticle towards PS4 while mutations in pxd cause the transformation of PS6 to PS5. The mutation also causes the appearance of ventral pits in the abdominal segments. These ventral pits are sensory organs present only in the thoracic segments. Pbx mutation causes a transformation of posterior T3 to T2 segment (Reviewed in Duncan, 1987).

Five mutations so far have been described that cause a gain of function in the *Ubx* domain, namely – contrabithorax1 (cbx ^{1}), haltere mimic (Hm), cbx^{2}, cbx^{3} and cbx^{Twt} (Bender *et. al.,* 1983, Bender *et. al.,* 1985). Cbx^{1} mutations show transformation of posterior compartment of wing to the posterior compartment of haltere disc (Morata 1975). A transformation of the distal region of wing to that of haltere is seen in Hm mutants (Bender *et. al.,* 1985). Cbx^{2} causes moderate but uniform transformation of the wing to haltere. However, only the posterior portion of the second thoracic leg is affected in this mutant (Bender *et. al.,* 1985). Cbx ^{3} causes transformation of anterior compartment of T2 wing and leg to T3 structures. cbx^{Twt} causes variable reduction of anterior dorsal T2 structure (Bender *et. al.,* 1983).

A study of these mutations has revealed a lot about the manner in which *Ubx* is able to impart identity to multiple segments. A comprehensive study of the *Ubx* locus in *Apis mellifera* is however not present. Molecular analysis of the Hox complex in honey bee shows similar organisation of Hox genes.

1.8: Choice of model organisms

There are multiple reasons for choosing *Apis mellifera* as a model system for the study so that a comparison of *Ubx* functions between honey bee and fruit fly can be made. The hymenopterans evolved 130 million years ago from ancestral

endopterygote lineage after diverging from other insects roughly about 250 million years ago (Porcelli *et. al.,* 2007). In comparison, the dipterans are the recently evolved insects. A comparison of the two insect orders in terms of the evolution of the CREs targeted by Ubx and the downstream functional targets would give an overview of the changes taking place over more than 250 million years.

Plate 1.9: Evolutionary relationships amongst various arthropods

An important reason for choice of the model organism lies in the fact that the technique used for identification of targets of Ubx relies on the availability of the complete genome sequence. *Apis mellifera* and a number of other hymenopteran genomes have been sequenced. The genome of *Apis mellifera* is the best annotated one amongst all the hymenopteran genomes.

A very important point that one looks at while selecting model organism for genomic studies is the availability of identical genomes amongst a population. While this is readily available in *Drosophila melanogaster*; in *Apis mellifera*, one can take the advantage of a single queen which is responsible for laying eggs in the hive and thus the entire progeny of hive can have identical genotype. Thus, by artificially inseminating the queen, one can achieve almost uniform genotype in the whole population which can be used for further studies.

Model organisms used in this study:

The current study is based on two model organisms – the common fruit fly and the European honey bee.

1.8.1:*Drosophila melanogaster*/ fruitfly:

Carl Freidrcik Fallen coined the term '*Drosophila*' meaning 'lover of dew' in 1823. Work on this model organism begun more than a hundred years ago and the first paper published using *Drosophila* as a model organism is ' The reactions of the pomacefly (*Drosophila amelophila* Loew) to light gravity and mechanical stimulation' by Fredrick W Carpenter in 1905. However, the first major work using *Drosophila* as a model was done by the group led by Thomas Hunt Morgan at

Columbia University starting around 1910. Since then, this small fly measuring around 7.5 mm in length has become one of the pivotal model organisms for understanding almost all aspects of Biology (reviewed by Roberts, 2006).

A number of advantages are associated with the use of fruit fly as a model for scientific studies, which include its short life cycle, absence of recombination in male flies, easy maintenance and handling of these flies. Over the period of time, biologists have developed a number of genetic tools like easy generation of transgenics in *Drosophila* with the help of P elements (Rubin *et.al.*, 1982); the UAS GAL4 system for targeted gene expression (Brad &Perimon, 1993); FLP- FRT system for somatic recombination (Xu & Rubin , 1993) ; knock out system (Gong &Golic, 2003) and a still more efficient and directed method of genomic engineering (Huang *et. al*., 2009).These genetic tools together with molecular biology techniques have made this organism one of the most important systems for biological studies.

As genomes of various eukaryotes were sequenced and the genes were annotated, it came to light that as many as 70% of genes of fruit flies are highly conserved with their human counter parts. This attracted even more interest in the already amenable model organism to study human diseases.

1.8.2: *Apis mellifera*/ The European honeybee:

The honey bee, *Apis mellifera* Linnaeus, is both an economically valuable species and a favourable one for genetic research. Discovery of parthenogenesis in honey bee by Dzierzon was an important advance in knowledge of reproduction. Count of yellow and black drones from unfertilized eggs of hybrid queen honey bee suggested Mendel the idea of counting and classifying the progeny of each mating which led to the famous laws of heredity. Earlier, Mendel and other biologists had tried to use honey bee as a model organism but due to lack of mating control techniques- the model never became popular. It was only when Lloyd R. Watson devised a technique for artificial insemination in this insect in the year 1926 that honey bees became a truly good model organism for various genetic studies. The insect is a compelling and very good model organism for studying animal behaviour and learning (Pankin *et. al*,

2001; Robinson, 1999). The insect is also a remarkable model for understanding the genetic basis of various sign languages (Johnson *et. al*, 2002; Von Frisch,1967). Developmental switches or polyphenism (Evans *et. al.*, 2000; Evans *et. al.* , 2001) are well known in this insect which provides a natural way of understanding epigenetics. There are two switches that particularly interest the researchers- longevity and fertility. While workers are nearly sterile, queen lays a huge number of eggs each year and lives 10-20 times longer than worker bees. Bees are also an important model for studying diseases. In honey bees and other social insects, one can study social elements in both transmission and progression of diseases (Traniello *et. al.*, 2002). Transgenesis in this model organism has been achieved rather recently (Robinson, 2000). With the availability of honey bee genome, a whole new field of comparative genomics has opened up, which allows us to look at all the above mentioned aspects in the light of evolution by comparing these traits in other organisms (where they have already been well studied).

1.9: Life Cycle

1.9.1: Life cycle of *Drosophila melanogaster*

One of the major advantage of using fruit fly as a model organism is its short life cycle, which allows one to have multiple generations in a relatively short span of time. The developmental period of *Drosophila* varies with temperature, the shortest developmental time being at 28 °C (7 days). At 25 °C, the development takes 9- 10 days while at 18°C it takes about 19 days.

Drosophila is a holometabolous insect – insects which have four stages in their life cycle- the embryo, larva, pupa and imago/ adult. The embryo are laid by females on any suitable food source, the embryonic development of the embryo takes around 24 hours at 25 °C. The embryo subsequently hatches into larva. There are three larval instars in the life cycle of *Drosophila*. The first instar larva feeds on the substrate on which it was laid and then moults into a second instar larva in around 24- 25 hours. The second instar larva continues to feed and in the next 24 hours moults into the next larval instar- third larval instar which is the longest larval instar. The third instar

too feeds but then slowly crawls upwards to a relatively drier area where it can pupate. The third instar moults into pupa after 30 hours. The pupal stage lasts for around 3- 4 days. During the pupal stage the larva is metamorphosing into the adult fly or imago and while doing so, it lyses most of the larval structures though some of the structures are maintained, especially the malphigian tubules, fat bodies, gonads and the nervous system. Most of the adult structures are formed from two sets of undifferentiated, mitotic cells that were carried by the larva during its development- imaginal discs (because they are for the imago) and the histoblasts.

Imaginal discs are small drop shaped packets of epithelial cells that ultimately form the epidermal structures of the adult like wings, eyes, mouthparts and genital ducts.

Histoblasts are the cells that are found in small nests within the larvae and form the abdominal epidermis and internal organs of the fly. They too grow by mitosis in the larval instars and differentiate during the pupal stage (Ashburner, 2005).

Plate 1.10: Life Cycle of *Drosophila melanogaster*

1.9.2: Life cycle of *Apis mellifera*

Life cycle of honey bee can be briefly summarised in the following table

Stage	Egg	L1	L2	L3	L4	L5	PP	P0 to P9	imago
Duration (days)	3	1	1	1	1	2	2	10	variable
Age (days)	0-3	4	5	6	7	8-9	11	12-21	
%	0-14	19	24	29	33	43	52	57-100	

Table 2: Absolute and relative duration of individual stages during the honeybee development L1 (L2, L3, …) larval stage 1 (2, 3,..); PP prepupa, P0 to P9 pupae at the age of 0 to 9 days.

(From PhD Thesis Stefan Eichmüller, Free University 1994)

In the first larval instar, the larva is slightly larger than the embryo, while in the second instar is twice the size of the embryo. In the third larval instar, the larva gets curved, but is located at the bottom of the comb. The fourth instar of the bee too is curved but fills the whole comb. The fifth instar on the other hand is called “streckmade” (stretching larva) and it is the first time that it is erected within the comb. Beyond this stage the nurse bees close the comb and pre pupa and puape are always found in a closed cell.

Plate 1.11: Life cycle of *Apis mellifera*

1.10: Early development in the two insects

1.10.1: Early development of fruit flies

Female fruit flies lay fertilised embryos, consisting of a nucleus, which undergoes mitosis every nine minutes. Till the twelfth nuclear division there is no formation of cell membrane around the nuclei, which results in the formation of a syncytium in which around 6000 nuclei are present in a common cytoplasm. However, after the completion of the nine divisions the nuclei move to the boundary to form a syncytial blastoderm. Membranes start forming after the twelfth division and the cellularization is complete at the end of fourteen divisions. Right at the syncytial stage, a small number of nuclei move towards the posterior end of the egg and form pole cells on cellularization and move towards outer side of the blastoderm. The blastoderm gives rise to the somatic tissue of the fly, whereas the pole cells ultimately give rise to the germline.

Plate 1.12: Cleavage in early *Drosophila melanogaster* embryo and formation of blastoderm.

Gastrulation in *Drosophila* makes it possible for all the future tissues to be derived from a single epithelial sheet. Gastrulation begins at three hours after fertilization and first involves the invagination of the ventral region to form a furrow along the vental mid line. The region forms meosderm in the future. Soon after this, the ventral ectoderm cells, which will give rise to nervous system leave the surface individually

and form a layer of neuroblasts between the mesoderm and ectoderm. Almost simultaneously, two tube like invaginations from at prospective anterior and posterior midgut which eventually form the endoderm of the midgut. No cell divisions take place during gastrulation but they begin soon after the process has been completed.
Plate 1.13: Gastrulation in *Drosophila melanogaster.*

Germ band extension too takes place during gastrulation during which the ventral blastoderm drives the posterior trunk regions round the posterior end onto what was earlier the dorsal side. The period is also characterised by the beginning of segmentation when equally spaced grooves define parasegments which are 14 in number- three contribute to mouthparts of the head, three to thoracic region and eight to abdomen.
Plate 1.14: Gastrulation, germ band extension and segmentation in *Drosophila melanogaster* embryo.

The larva hatches around 24 hours after fertilisation and the larva undergoes two moulting cycles to reach the third instar larval stage after which pupation takes place. The adult flies emerge out of the pupa.

1.10.2: Early development of the honey bee

Very few studies have been done on the embryogenesis of honey bee. Some of them have been done very early in the 20^{th} century based on observations under a light microscope (Nelson, 1915), while few have been done in the later half of the 20^{th} century (DuPraw, 1967) (one of them being based on SEM studies of the embryo from ovipostion to hatching). Honey bee embryogenesis does not show as specialised traits as *Drosophila* (especially the head involution seen in fruit flies) but conforms more to the general scheme of insect development.
The fertilised nucleus in the embryo divides till the 10^{th} mitotic cycle in a syncytial cytoplasm. The nuclei then enter the periplasm where they start the next mitotic cycles. The blastoderm takes a two layered appearance after the 14^{th} mitotic cycle

(about 10 hours after oviposition). Dorsal strip starts forming around 24 hours after oviposition. In this process the lateral and dorsal blastoderm parts move somewhat on the ventral side

Gastrulation takes place from around 33 hours AEL to 40 hours AEL. During the process first the germ layers get segregated (process starting 33 hours AEL). A pair of slightly curved furrows becomes visible on the either side of the ventral midline. The blastoderm is divided into three parts by the extension of these furrows in the posterior direction: the lateral plates (ectoderm); the middle plate (mesoderm) and two patches of endoderm anterior and posterior to the mesoderm. As the furrows extend towards the posterior pole, the mesoderm sinks inwards while the prospective mesoderm and ectoderm start disconnecting along their borders. The ectodermal plates move medially over the mesoderm where they start fusing in the differentiation center (DC). The fusion of ectoderm continues anteriorly and later posteriorly which internalises the entire endoderm. The completion of fusion of ectodermal plates completes gastrulation. At the posterior end, some globular cells become visible at the end of gastulation, which become germ cells in the future.

Body segmentation starts soon after the completion of gastrulation (34 hours AEL). Some transverse grooves start to extend across the ectoderm around the differentiation centre. Very soon further transverse grooves form both in the prospective ectoderm and the mesoderm. Every second segmental groove fades rather suddenly at around 40 hours AEL, so that a pattern of paired segments result in early germ band.

Plate 1.15: Blastoderm, gastulation and early germ band stage of *Apis mellifera.*

In the germ band stage, which begins at 40 hours AEL and continues till 55 hours AEL, the perivitellinespace reach their maximum extent. During this stage, serosa, amnion, midgut, nervous system and other ectodermal structures are formed.

Finally, dorsal closure takes place during which the more lateral cells of the amnion start moving towards the dorsal midline, partly pushing over each other. This process begins 55 hours AEL when the organ formation in embryo is quite advanced. At the

end of this process both lateral ends of ectoderm meet dorsally and fuse; the mesoderm fuses only around 5- 6 hours before hatching of the embryo.

Plate 1.16: Advanced germ band stage and dorsal closure.

The embryo hatches 3 days after oviposition and then continues into five larval instars, the first four lasting for 1 day each while the fifth one lasts for 2 days. The larva then forms a pre pupa, which lasts for 2 days. The prepupa ultimately gives rise to pupa that lasts for 10 days when the adult emerges out (Fleig, 1986).

1.11:Axis formation in the early embryos

Essentially, three axes can be identified in the adult insects

1. Anterior – posterior axis
2. Dorsal- ventral axis
3. Proximal- distal axis

Two of these axes; the anterior- posterior axis and dorsal – ventral axis form during embryogenesis while the proximal distal axis is somewhat a secondary axis that is formed during patterning of adult organs.

1.11.1: Formation of anterior- posterior axis

1.11.1.1:*Drosophila melanogaster*

Morphogen gradients are established in the *Drosophila* embryo along the anterior-posterior and dorsal ventral axis. The gradient essentially is established by proteins and mRNAs, which are differentially distributed along the cytoplasm of the embryo. Anterior-posterior axis gradient is established soon after fertilization while the dorsal – ventral gradient is established a bit later.

1.11.1.1a: Maternal genes

Asymmetry is imposed on the *Drosophila* oocyte during oogenesis. The cytoplasmic connections between the nurse cells and the oocyte are made at one end of the oocyte and this end forms the anterior end of the embryo. Maternal genes play an important

role in early development of *Drosophila* embryo and are also responsible for setting up morphogen gradients.

The anterior system, which is responsible for the development of head and thorax, uses maternal germline products namely Exuperantia, Staufen and Swallow to localise *Bicoid (Bcd)* mRNA (encoded by maternal genes) at the anterior end. *Bcd* mRNA is transcribed in the nurse cells and transported to the oocyte. When translated, the gradient of Bcd is produced by the diffusion of the protein product from the localised source at the anterior end.

The posterior system gives shape to the posterior end of the embryo and various gene products act to localise *nanos*. Cappacunio and Spire act to localise Staufen at the posterior pole which in turn localises *Oskar* mRNA. The complex of Staufen and *Oskar* RNA act to localise Vasa, an RNA binding protein. The posterior pole contains two morphogens- Nanos controls the abdominal development and Oskar controls the formation of germ cells. Nanos protein acts to prevent the transcription of hunchback (Sprenger *et. al*., 1992 &Struhl *et. al*., 1992).

1.11.1.1 b: Gap genes

Bicoid and Nanos gradients determine the expression of next set of genes called gap genes. The four gap genes- *Hunchback (Hb), Kruppel (Kr), Giant* and *Knirps* respond to Bcd gradient and also respond to each other. The synthesis of *Hb* mRNA is regulated by Bcd and the gradient of Hb in turn regulates the expression of other genes. Hb expresses itself in a broad anterior zone , with a gradient decline in the middle of the embryo. High levels of Hb repress *Kr*, and as Hb expression falls, the expression of Kr rises. Some level of Hb is needed for *Kr* expression, so when the level of Hunchback falls, the level of Kr too falls. *Giant* too responds to Hb in a similar way and *Knirps* expression begins when the Hb is absent (Simpson-brose *et. al.*,1994).

1.11.1.1c: Pair rule genes

All the gap genes, besides regulating one another, regulate another class of genes called pair rule genes. Each pair rule protein is found in a pattern of seven stripes that mark the future parasegemnts (Lawrence *et. al.*,1989)

1.11.1.1d: Segment polarity genes

The pair rule genes in turn regulate the expression of segment polarity genes, which are expressed in 14 stripes. Each of the 14 segments is divided into an anterior and posterior compartment by the activity of segment polarity genes. Various types of proteins like transcription factors, transmembrane proteins, kinases, cytoskeletal proteins and morphogens are encoded by these segment polarity genes. The cells that define the boundaries of anterior and posterior compartment express *wingless (wg)* and *engrailed (en)* in a reciprocal manner. Wg protein is secreted from a cell and acts upon a cell on its posterior side. This causes En protein to be expressed which leads to the production of Hedegehog (Hh) protein, which is secreted. Hh acts on its anterior cells to maintain *wg* expression. This circuit defines the boundaries between anterior and posterior compartment in a segment (Kornberg, 1981; Kornberg *et. al.,*1985 & Tabata *et. al.,* 1995)

1.11.1.1e: Homeotic genes

The final stage in the specification of the segmented body plan begins with the expression of homeotic genes. Initiation of homeotic selector gene expression requires input from gap gene proteins and the pair rule proteins that control their expression within parasegmental domains. Gap and pair rule genes acts via cis regulatory elements to control the expression of homeotic genes. The expression of homeotic genes are also regulated by one another. In general, the posterior genes, inhibit the expression of more anterior genes in the posterior part of the embryo. As the expression of gap and pair rule proteins is transient, other mechanisms are used in the later stages of development to ensure the correct spatio-temporal expression of homeotic genes. Polycomb and trithorax group of genes act in opposing manner to stabilise homeotic gene expression. By remodelling chromatin structure to more compact state, the proteins of polycomb group inactive cis-regulatory regions that control the homeotic gene expression. On the other hand the trithorax group maintains the chromatin in an open state that favours their transcription (Lewis, 1978; Beachy *et. al.,* 1985; Karch *et. al.,*1985 & Martin *et. al.,* 1995).

Besides the anterior and posterior system, there is a terminal system that is responsible for the development of specialized structures at the unsegmented ends of the embryo. The system depends on maternal somatic genes that create asymmetry in the follicle cells. They lead to localised activation of the transmembrane receptor coded by torso, which ultimately shapes the acron at the head and telson at the tail.

1.11.1.2:*Apis mellifera*

Most of the studies on axis formation in honey bees rely on analysing the conservation and divergence of the axis specifying genes, when compared to *Drosophila*. In honey bees , the gene *Bcd* is missing implying that other genes must carry out its function in its absence like *Orthodenticle* and *Hb* as has been shown in *Tribolium* (Schroder *et. al.*, 2003) and *Nasonia* (Pultz*et al*, 2005).

Honey bee genome has orthologs of almost all the major gap genes, pair rule genes and the segment polarity genes.

The expression of segment polarity genes have been conserved between honeybees and *Drosophila*. The depletion of *en* by RNAi has also shown that engrailed plays an important role in parasegment boundary formation (Beye *et. al*., 2002).

Honey bees contain all the ten Hox genes. The honey bee Hox gene cluster is not split into two as in *Drosophila*; and all the genes are transcribed from the same strand indicating lack of inversions in this insect. The Hox complex lies on chromosome 16 and is around 1.37 Mb in length (Dearden *et.al*., 2006). The size of Hox complex in *Drosophila* is 0.66 Mb (Drysdale and Crossby, 2005), in *Tribolium* it is 0.7 Mb in length (beetlebase); in *Schistocerca* it is around 0.7 Mb in length (Ferrier and Akam, 1996) and in *Anophles* it is 1.18Mb in length. The large size of the complex may be explained by large intergenic regions present as compared to other insects. The collinear expression of hox genes too is conserved in honey bees. (Dearden *et. al.*,2006).

1.11.2: Formation of dorsal- ventral axis

1.11.2.1: Dorsal- ventral axis in *Drosophila melanogaster*

During the syncytial blastoderm stage the fly embryo is divided into three primary domains along the DV axis. These include a ventral domain, which gives rise to mesoderm; a ventro lateral domain which gives rise to ventral epidermis and central nervous system; and a dorsal domain which gives rise to dorsal epidermis and amnioserosa.

Drosophila embryo is subdivided along its D-V axis by two opposing gradients- gradient of Dorsal and Decapenptaplegic (Dpp). The concentration of Dorsal is maximal in the nuclei of the ventral region of the embryo and decreases dorsally; while Dpp concentration is maximal at the dorsal region and its concentration decreases ventrally. The use of Dorsal to pattern the D-V axis is unique to invertebrates but the Dpp pathway is conserved in both invertebrates and vertebrates.

Dorsal nuclear concentration gradient

The movement of oocyte nucleus from an initial posterior position to an anterior-dorsal location initiates the beginning of establishment of dorsal-ventral polarity.

Accumulation of high levels of *Gurken* mRNA between the nucleus and plasma membrane followed by a second round of Gurken signalling by oocyte results in a dorsal to ventral gradient of *Epidermal growth factor receptor (EGFR)* activation in the follicle cell epithelium. EGFR signalling in follicle cells repress Pipe expression; as a result of which *Pipe* expression is restricted in the follicle cells along the ventral side of the embryo. Pipe modifies an unknown proteoglycan substrate and deposits it in the space outside the oocyte before the vitelline membrane and the chorion are secreted. This proteoglycan activates Spatzle in the ventral side of the embryo, which in turn activates Toll- which is distributed uniformly in the embryo but gets activated only in the ventral side due to the localised activation of its ligand. The spatially restricted activation of Toll leads to graded nuclear uptake of Dorsal. Dorsal is encoded by maternally supplied mRNA, which is present uniformly in the embryo. Toll, which is activated ventrally enables a graded nuclear import of the protein- thereby establishing its gradient. (reviewed in Wylie *et. al.*,1996)

Dorsal in turn modultes about 50 to 60 genes in the blastoderm embryo. Dorsal regulates these genes via enhancer or silencer elements. The binding sites of these genes decide the threshold concentration of Dorsal required for their activation or repression. *Twist* and *Snail* are only activated at high concentration of Dorsal present in the ventral region of the embryo; whereas *Short-gastrulation (Sog)* and *Rhomboid (Rho)* are activated at relatively lower concentration of Dorsal as well. Essentially, by modulating the expression of various genes, Dorsal imparts a ventral fate to the fly embryo (reviewed in Belvin *et. al.*, 1996).

Dpp concentration gradient

Dpp is transcribed almost uniformly in the embryo except in the ventral region where its transcription is inhibited by Dorsal. Patterning of the dorsal ectoderm requires another ligand called Screw. Dpp and Screw function as dimers to impart a dorsal fate to the embryo. Punt, Thickveins and Saxophone are the kinases, which are essential for essential for signalling by Dpp and Screw. Redistribution of Dpp is dependent on Sog. Sog forms a gradient with a high concentration at the dorsal edge of Sog expression and lower concentrations towards the dorsal midline. Sog binds to both Dpp homodimers and Dpp/Screw heterodimers in complexes that also contain Twisted gastrulation . When present in this complex Dpp and screw are unable to bind to their receptors. This prevents them from interfering in the normal developmental program in the neuroectodermal and mesodermal domains. At the same time Sog facilitates the diffusion of these ligands through the perivitelline space to the dorsal midline, resulting in zone of high Dpp concentration at the dorsal midline.

Dpp, therefore acts a morphogen and directly affects the phenotype of the cells, the cells that receive the maximum concentration of Dpp assume a dorsal fate and as the concentration of dpp decreases the cells start assuming a ventral fate. (De Robertis *et. al.*, 1996).

1.11.2.2: Dorsal- ventral axis in *Apis mellifera*

The genes required for dorsal- ventral patterning are conserved in honeybees except that *Gurken* is absent in honeybees. Also the expression of few genes like *Toll* is slightly different in honeybees than in *Drosophila*. These results indicate that there may be a slightly different mode of defining the dorsal- ventral axis in honeybees (Dearden *et. al*., 2006).

1.12: Differences in morphology of the dorsal second and third thoracic organ in two insect species (plate 1.17)

Morphological differences between wing and haltere in *Drosphila melanogaster*

Wing of the common fruit fly, like any other insect is an organ for flight while the haltere in these flies is a specialised organ that is used by the insects as a balancing organ. The wing is a flattened structure composed of veins and interveins. Veins are hollow structures through which the trachea and the nerve fibres run. The intervein region is formed by apposition of two layers and in the adult fly the cells in this region are dead. Haltere, On the other hand is a bulbous structure and is much smaller than the wing. It is devoid of any vein intervein pattern and serves to detect the flow of air during the flight of fruit flies.

Morphological differences between fore and hind wing of *Apis mellifera*

The fore and hind wing of honey bee are almost identical except for few differences. Unlike haltere of fruit fly which is very different from the wing; the hind wing of honey bee is only slightly smaller than the fore wing. There are also subtle differences in the venation pattern between the fore and hind wing. A important difference that is seen in the fore and hind wing is the orientation of bristles on the edge of the wings. A set or interlocking bristles are found in the posterior end of the fore wing while the complementary set is found in the anterior end of the hind wing.

This interlocking mechanism serves to lock the fore and hind wing of honey bee during flight.

1.13: Development of wing in the fruit fly

Imaginal discs are the sets of undifferentiated, mitotic cells that are carried by larva of fruit fly during development. The wing and haltere (the dorsal second and third thoracic appendage of *Drosphila*) are formed from two such imaginal discs- the wing and haltere imaginal discs.

Two organizing centres located at the boundary of antero posterior and dorso ventral compartments are responsible for shaping the wing blade. The wing imaginal disc consists of about 20 cells when it is formed during embryonic development. These cells divide and ultimately the wing imaginal disc in the late third instar larva consists of 75000 cells. The wing and leg discs of *Drosophila melanogaster* have a common precursor which later separate as result of dorsal segregation of the wing disc. The common precursor is established at the A-P boundary within the mesothroacic segment and consists of *Engrailed (En)* expressing and non expressing cells. It therefore seems that the A- P boundary is inherited from the embryos and is maintained in the larvae. The common disc precursor also includes *wg* expressing cells but the wing disc does not inherit any of the *wg* expressing cells after the segregation and hence has no obvious D-V boundary (Reviewed by Klein , 2001 and Brook *et. al*., 2006).

The mature late third instar larval disc is a flattened sac with two distinct surfaces- a thin peripodial membrane and a thicker folded disc epithelium. The disc epithelium makes most of the wing blade and hinge and also the body of the adult fly. During pupation, the wing disc everts, folding upon itself to form the apposed dorsal and ventral epithelia of the wing blade. Once the wing disc everts, the basal sides of the dorsal and ventral surface of wing epithelia come together and proveins from as broad gaps or lacunae. Approximately 6-8 hours APF, the wing secretes an apical cuticle. Soon after this, the wing inflates and the dorsal and the ventral surfaces are again apart. The vein morphology reappears by the apposition of the dorsal and ventral surfaces around 18-30 hours APF. The pupal cuticle detaches itself from the

apical surface and final adult cuticle is secreted at around 36 hours APF (Reviewed by Blair, 2007) (**Plates 1.18& 1.19**).

En and its homologue *Invected* impart identity to the posterior compartment of the wing disc. The cells that comprise the posterior part of the common disc precursor inherit *en* expression from embryonic primordium (Vincent *et. al.*, 1992). Cells in the posterior compartment communicate with the adjoining cells in the anterior compartment by through localized expression of secreted molecule called Hh (Lee *et. al*., 1992). Hh activity in posterior cells is required for development of both the compartments in imaginal discs. Hh can induce patterning only in the anterior compartment even though it is expressed in the entire posterior compartment. It induces expression of *Dpp* in a narrow stripe of cells near the compartment boundary (A-P) (Basler *et. al*., 1994). Dpp in turn organises patterning and controls growth symmetrically in both the compartments. Hh also turns on *patched (ptc), knot (kn)* and *en*. Dpp acts as a morphogen and turns on *vestigial (vg), optomoter blind (omb)* and *spalt (sal)* at successively higher thresholds. These genes are responsible for patterning wing blade in various ways along the A- P axis (**plate 1.20**).

The DV boundary too serves as an organizing center that controls growth and specifies spatial pattern along the dorsal ventral axis. The wing disc that originates from common disc precursor lacks *wg* expressing cells and hence has no obvious D-V boundary. However, a defined *Apterous* (*Ap*) expression has been observed during the first instar larval stage. *Vein (Vn),* a component of EGFR pathway has a role to play in the establishment of *ap* domain in the wing disc. *Vn*, itself is expressed in the dorsal part of the wing disc and diffuses ventrally to generate a gradient of EGFR activity that provides two thresholds for gene activation- a high threshold maintains its own expression while a lower threshold is needed to activate *Ap* expression (Wang *et. al*., 2000).

Wg starts expressing in the second larval instar in the wing disc by the activity of *Hh* signalling pathway. *Wg* is expressed more in the ventral part of the disc and it suppresses the activity of *Vn,* thereby restricting the activity of *Vn* to the dorsal part of the wing imaginal disc. The EGFR pathway suppresses *Wg* in the dorsal part of the

wing and if interrupted, allows the expression of *Wg* in the dorsal part of the wing disc. This antagonistic relationship between EGFR pathway and *Wg* allows the wing disc to be divided into pouch and notum regions. The region expressing *Ap* adopts a dorsal fate while the part of disc not expressing it remains as the ventral region. Interactions between dorsal and ventral cells are necessary for establishing the organising center. *Fringe (Fng)* and *Serrate (Ser)* are the genes that mediate signal from dorsal to ventral cells (Thomas *et. al.*, 1991 & Irvine *et. al*., 1994) *Ser* encodes a membrane spanning protein with extracellular EGF repeats while Fringe encodes a secreted protein. *Ser* is activated and kept on by *Ap* in the dorsal compartment in the second larval instar. In the mid – late third larval instar, *Ser* fades from much of the dorsal side and intensifies at the dorso- ventral boundary. *Delta (Dl)* essentially is the gene that mediates signal from ventral to dorsal side. It is transcribed throughout the disc as early as mid second instar larva but is expressed more strongly ventrally. In the mature wing discs, *Dl* is expressed as two bands that flank the *Wg* on D-V stripe.

Dl and Ser serve as the ligands of Notch and activate their receptor at the DV boundary (Rebay *et. al.*, 1991). Activation of Notch at the boundary leads to activation of *Wg* and *Cut (Ct)* at the margin. Wg acts a morphogen and diffuses in either direction of the D- V boundary to shape the wing blade. Wg also refines its expression and promotes the expression of *Dl* and *Ser* on either side of its expression domain to create a positive feedback loop that maintains Notch signalling and *Wg* expression. Expression of *Ct* gene product at the margin is responsible for maintenance of *Wg* expression as well as preventing the *Wg* target genes from responding to it at the margin. *Wg* targets *Vg, Distalless (Dll), Achaete (Ach)* at increasingly higher thresholds (**plates 1.20 & 1.21**).

Modification of wing to haltere in the third thoracic segment of *Drosophila*

Haltere is a balancing organ found in all dipterans in the third thoracic segment by modification of the wing fate, effected by Ubx. Ubx functions at multiple levels in hierarchy of wing disc development and promotes the formation of halteres by suppression of various wing patterning genes (Weatherbee *et. al*., 1998; Shashidhara

et. al., 1999; Galant *et. al*., 2002; Mohit *et. al*., 2003). Various components of important signalling pathways that are responsible for shaping the wing (Dpp, wnt, Notch, EGFR and Hedgehog signalling pathway) have been found to be targets of Ubx by microarray studies and also in studies that have focussed on understanding the role of these pathways in shaping the halteres.

A recent study that has been carried out to identify the targets of Ubx in haltere at various stages of development suggested that almost an equal number of genes are up regulated and also down regulated in the haltere as compared to the wing (Pavlopoulos *et. al*.,2011). Recent studies have also been carried out to identify the direct targets of Ubx in halteres to understand the general mechanism of gene selection by Ubx in halters (Agrawal *et.al*., 2011; Choo *et.al*., 2011, Slattery *et. al*., 2011).

1.14: Development of wing in the honey bees

Wing development in honey bees *(Apis mellifera)* has not been published earlier in literature. However, a very good amount of work has been done and reported on the wing development of various hymenopteran insects like ants. Wing patterning network has remained conserved across holometabolous insects for 300 million years (Carroll, 1994). The expression of these wing patterning genes has been studied in various castes of ants. Ants show a considerable degree of polyphenism with worker and solider castes having small or no wings. The expression of the wing patterning genes, though interrupted in various castes has remained conserved in all castes of ants (Bowsher *et.al*., 2007; Abouheif *et.al*., 2008). Few studies have shown that the wings of few castes of ants may be smaller in size but the expression of wing patterning genes is quite identical in these smaller wings when compared to that of other winged castes (Bowsher *et.al*., 2007). The difference between the fore and hind wing of honey bee has been described and it was expected based on other studies that the expression of patterning genes would be identical and conserved in the fore and hind wing of honey bee. The expression pattern of these genes has been discussed later.

Objectives

In order to understand the roles played by regulatory elements in shaping the haltere identity in *Drosophila* and shaping a smaller hind wing in honey bees, this study tried to identify the direct binding sites of Ubx in the two insects, compare them and subsequently identify the mechanisms that shape haltere instead of a hind wing in dipterans (or how Ubx essentially imparts diversity in different insects).

The objectives of the work can be summarised as:

1. Identification of direct binding regions of Ubx in the honeybee, *Apis mellifera.*
2. Identification of direct target genes of Ubx in hind wing disc of *Apis mellifera* and compare the same to those in halteres of *Drosophila melanogaster* and identify the diptera specific and the hymenoptera specific targets that might be involved in shaping the haltere or hind wing respectively.
3. Analysis of the direct binding regions using bioinformatics tools in order to understand the mechanisms that have enabled Ubx to shape a haltere in fruit flies instead of a hind wing in honey bee.
4. Analysis of the genes that are Ubx targets in the two insects and understanding the probable role of diptera specific Ubx target genes in shaping the haltere.

Plates

Chapter 1

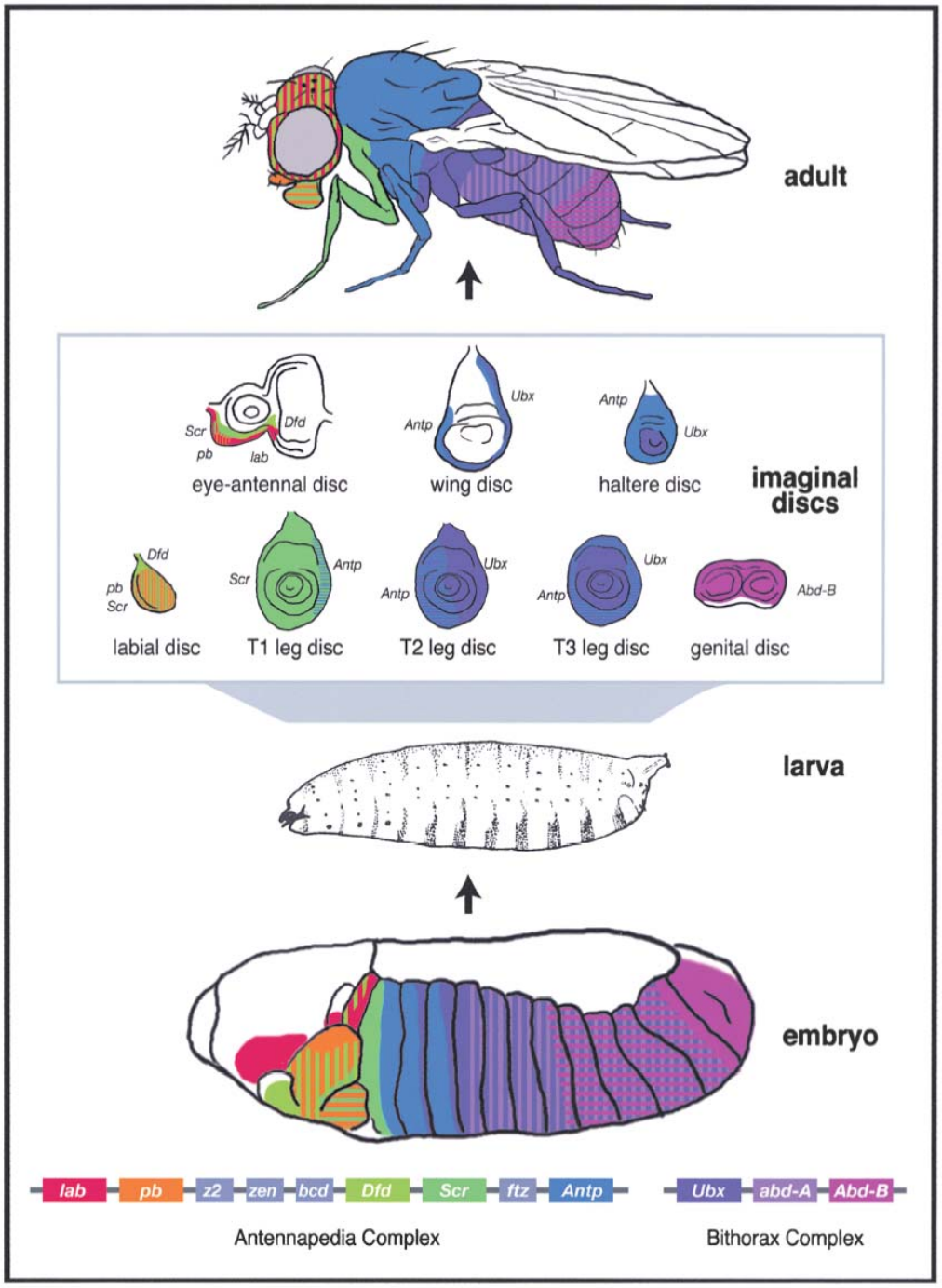


Plate 1.1 : Hox genes in *Drosophila melanogaster* (Courtesy: Hughes *et.al.*, 2002)

Hox genes are expressed along the anteroposterior axis in fruit flies and impart specific segmental identity to the segments in which they express.

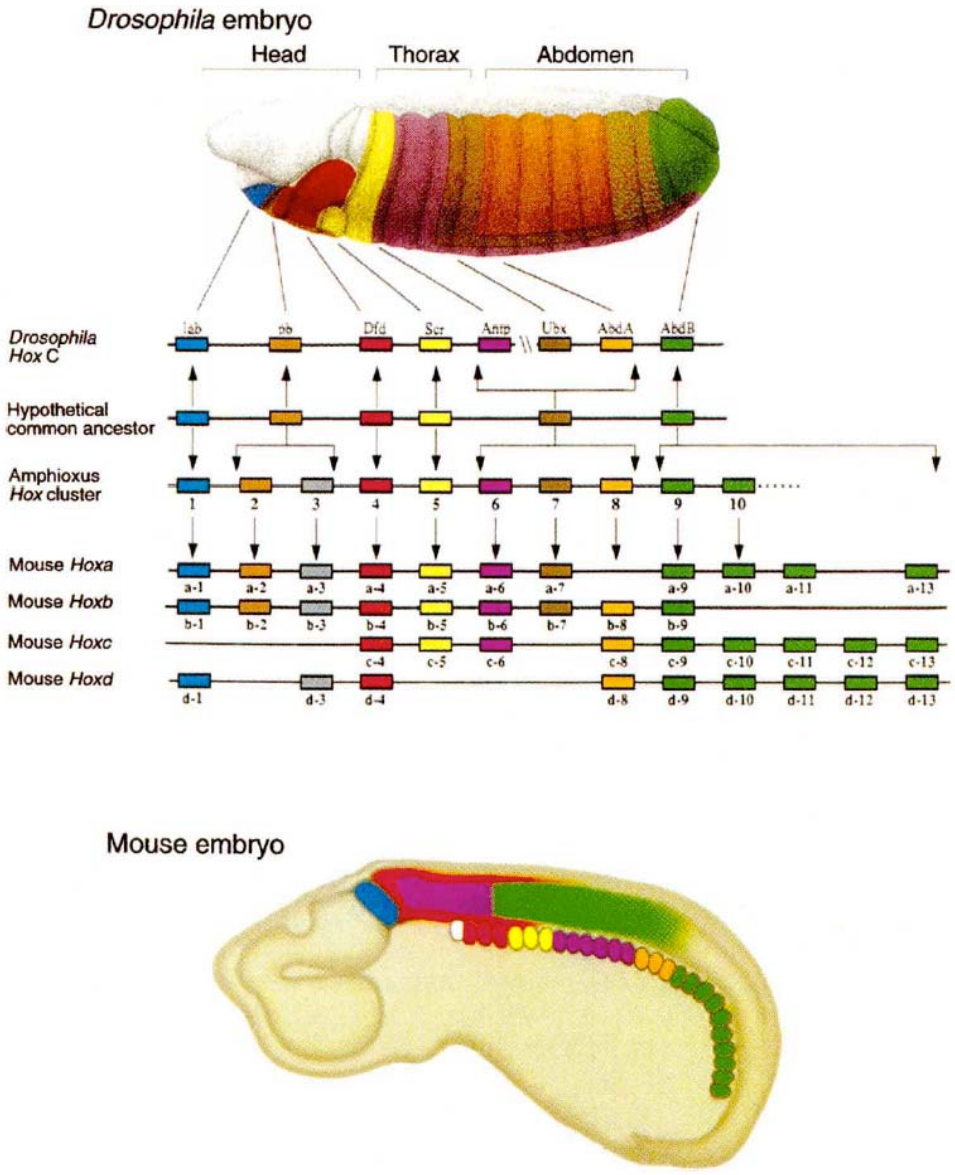


Plate 1.2: Conserved Hox gene organisation and expression across various organisms (Courtesy: Carroll, 1995)

Both fruit fly and mouse Hox gene expression correspond to the gene order within the hox complex. The evolutionary relationship between mouse amphioxus, fruit fly and ancestal hox complex is shown in the middle.

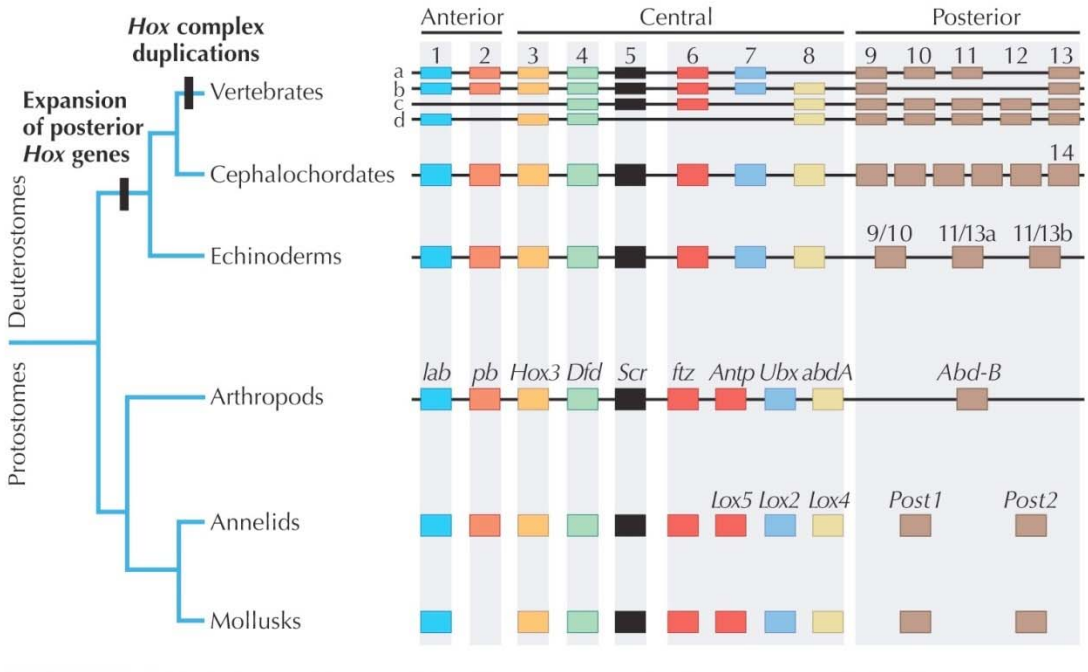


Plate 1.3: Evolution of Hox gene complex in metazoans (Courtesy: Patel)

Expansion of Hox complex seems to have occurred in multiple lineages. However, it seems that the common ancestor of Protosome and deuterosome had a hox complex of at least eight genes. The diversitifcation in body plan in higher vertebrates seem to be co related with and expansion in number of Hox genes but all the arthorpods share the same number and type of Hox genes.

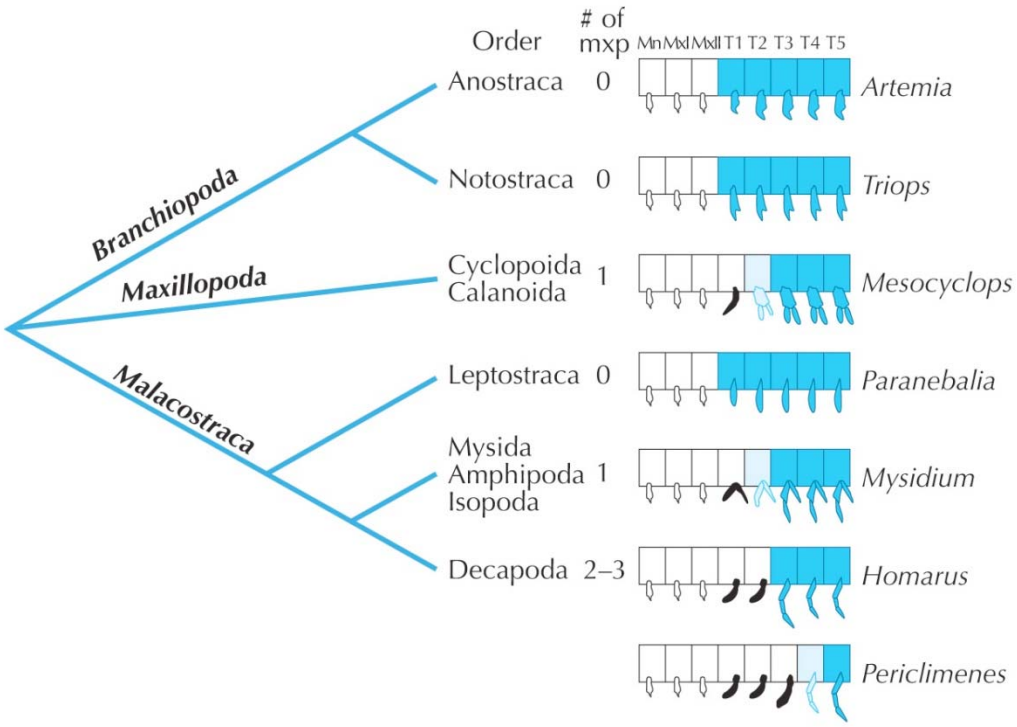


Plate 1.4: Segmental specialization and *Ubx* expression amongst various chelicerates (Adapted from Averof *et. al.*, 2007).

Dark blue shows high *Ubx* expression while light blue shows low *Ubx* expression. White is absence of *Ubx* expression. The limbs shown in black are the maxillipeds. Early embryonic pattern co relates with the morphology of appendages. Ubx is produced in segments that will produce locomotory organs of thorax but not in maxillipeds of thorax.

Change in maxilliped number co relates with the shift in expression pattern of *Ubx.*

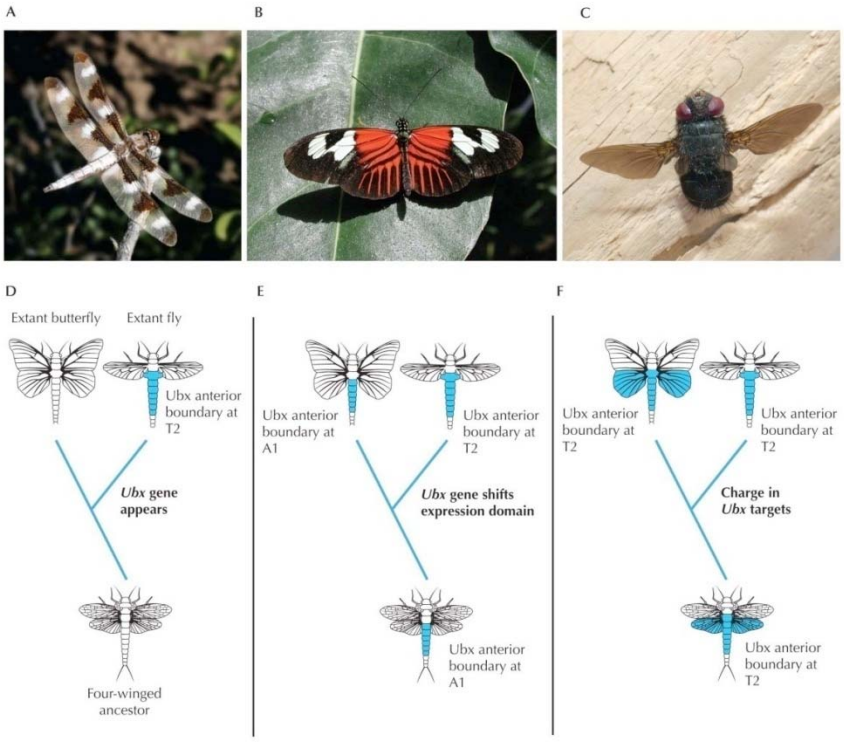


Plate 1.5: Mechanisms which may be used by evolution to modify the four winged ancestral insects to two winged dipterans. (courtesy: Patel)

A: Four winged ancestral insects (dragon fly belonging to the order Odonata). The fore and hind wings are similar in shape, size and ornamentation.

B: Four winged butterflies (belonging to more recently evolved order Lepidoptera). The hind wings are slightly smaller than the fore wing and different in ornamentation.

C: Two winged dipterans. The common fruit flies have reduced bulbous structure called halters instead of hind wings.

Three different hypotheses for the modification of wing into haltere have been proposed, of which the third (F) seems to be the most appealing. (At least in insect order, the expression pattern of Ubx and its function seems to have been conserved).

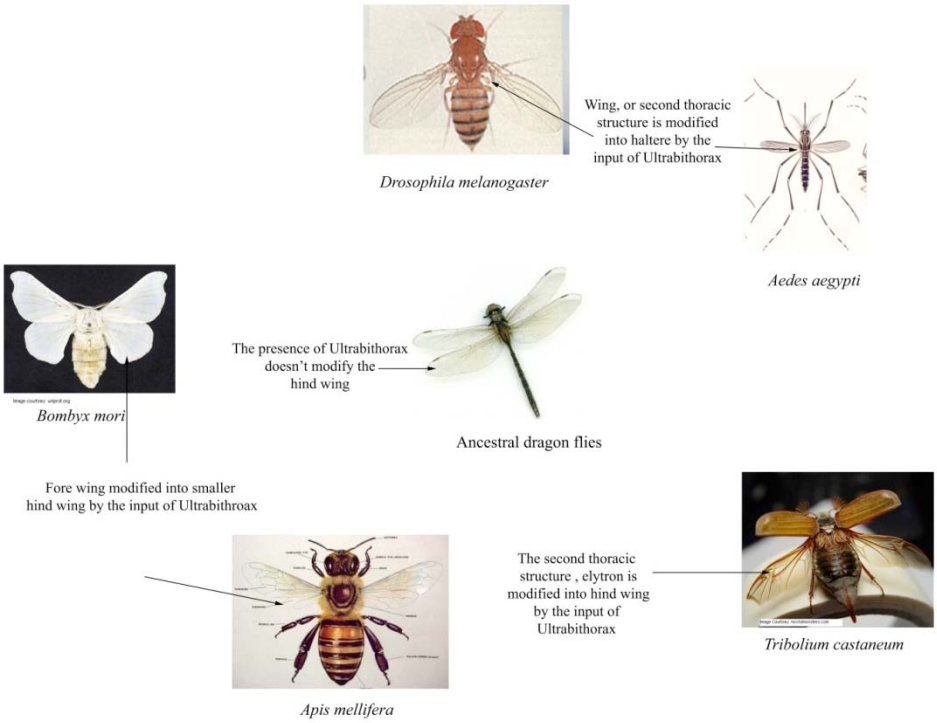


Plate 1.6: Modifications of wing (dorsal second thoracic structure) to various degrees in different insects

The picture shows ancestral insect (dragon flies) which have identical fore and hind wing (Ultrabithorax is expressed in the hind wing of dragon flies). In other insects like silkworm and honey bee, it can be observed that the hind wing is slightly smaller than the fore wing while in dipterans like mosquito and fruit flies, the fore wing is modified to a smaller haltere. Completely different transformation takes place in beetles, where a elytron is present in the second thoracic segment and it is transformed into a wing in the T3 segment.

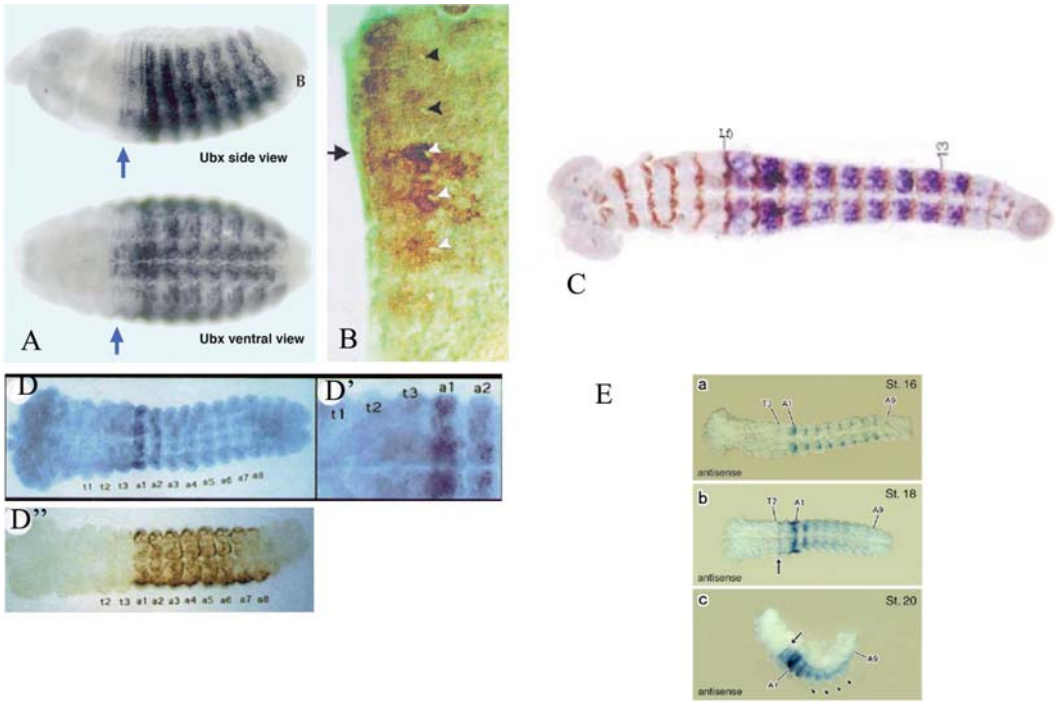


Plate 1.7: RNA in situ and antibody staining on various insect embryos showing localisation of *Ubx* transcripts and proteins

A. Localisation of Ubx protein in *Drosophila* embryo (adapted from Lawrence, 1992). Arrow indicated boundary between T2 and T3 segment.

B. Ubx/ abdA expression in *Apis mellifera* embryos (adapted from Walldorf *et. al.*, 2000). Arrow indicated the boundary between thorax and abdomen while pits are indicated by triangles

C. *Ubx* transcripts in *Tribolium* embryos that have completed germband stage (blue). Engrailed staining is shown in red (adapted from Bennett *et. al.*, 1999). Ubx is detected from PS 5- Ps 16, the expression being maximum in PS6.

D. D' and D'' show the distribution of *Ubx* transcripits in early *Precis coenia* embryos. Maximum expression is seen in A1 while diminished expression is seen in posterior compartments. Antibody staining on early embryos shows

the protein expression in entire anterior embryo and slight Ubx expression in the lateral part of T2 and T3 segments. (Adapted from Warren *et. al*., 1994).

E. Expression pattern of *Ubx* in *Bomby mori* embryos. Embryos are shown at three different stages (stage 16 in a, stage 18 in b and stage 20 in c). While highest level of Ubx is detected in A1, lower levels are seen in T3 and A9.

(Courtesy: Masumoto *et. al.*, 2009)

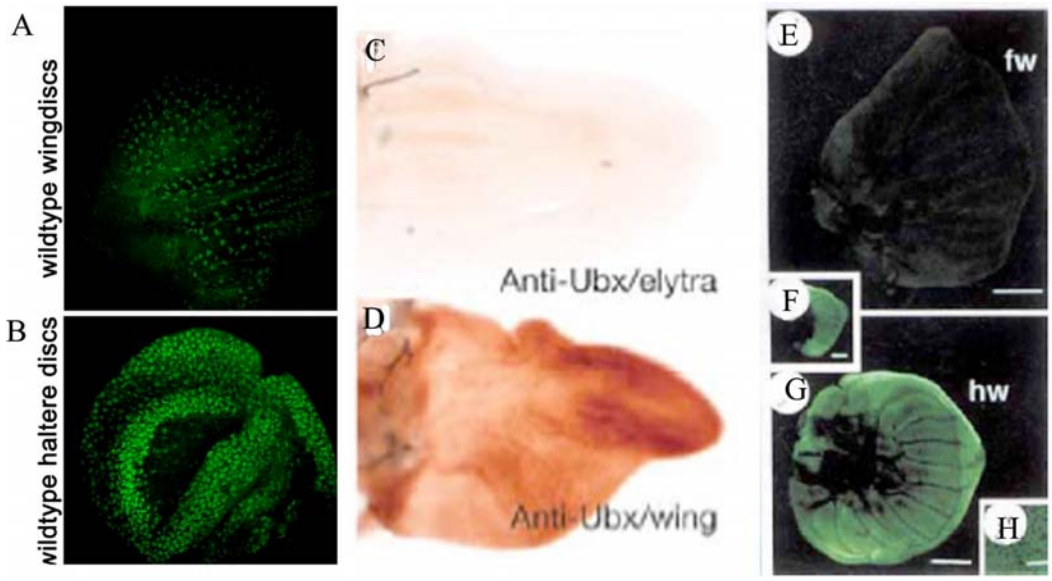


Plate 1.8: Expression of *Ubx* in dorsal second and third thoracic structures of various insects.

A: *Ubx* is expressed only in the peripodial membrane of *Drosophila* wing imaginal discs and is absent from disc proper cells

B: *Ubx* is expressed ubiquitously in the whole haltere disc.

C. In the elytra of beetles (the second thoracic structure) Ubx is almost absent.

D. In the beetle wings (the third thoracic structure) Ubx is highly expressed.

E. Ubx is absent from the fore wing disc proper cells of *Precis coenia*

F, G, H. Ubx is expressed in the hind wing of *Precis coenia* in the fourth (F) and fifth instar

(Courtesy: Agrawal *et. al*., 2011 ; Tomoyasu *et. al*., 2005 & Warren *et. al*.1994)

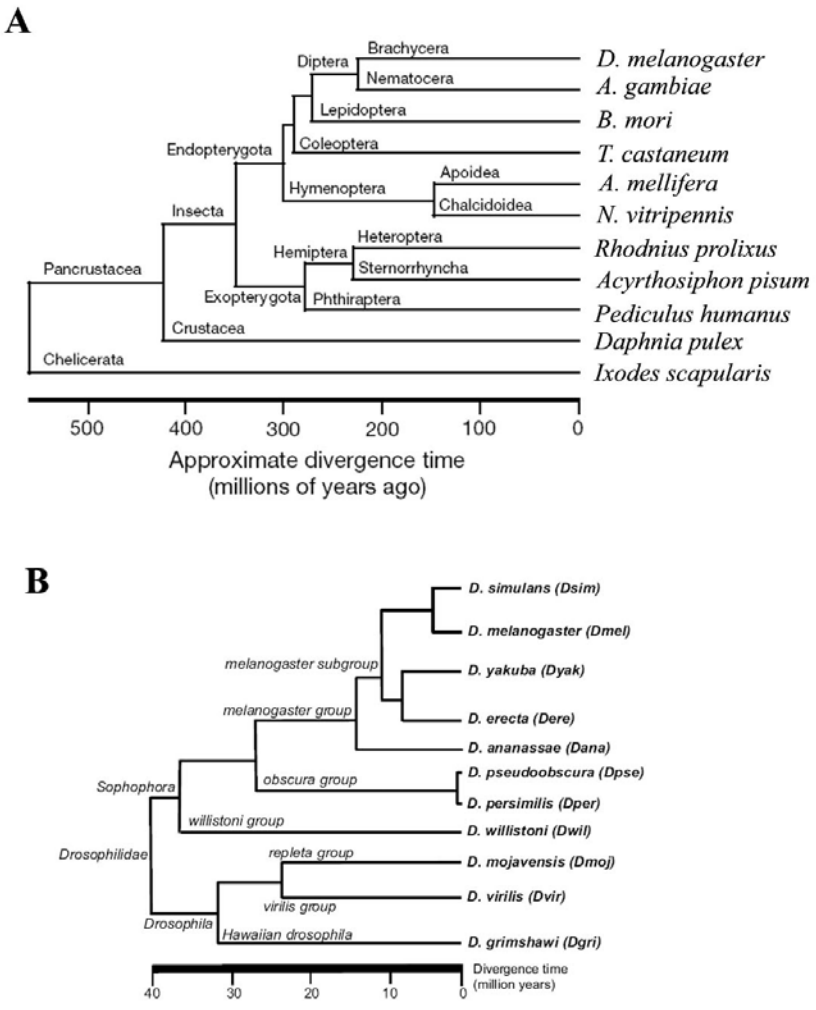


Plate 1.9: Evolutionary relationships amongst various arthropods

A. Divergence time between insects and other related arthorpods (adapted from Hony bee genome sequencing consortia , 2006)
B. Evolutionary history of various Drosophila species (Porcelli *et. al.*, 2007)

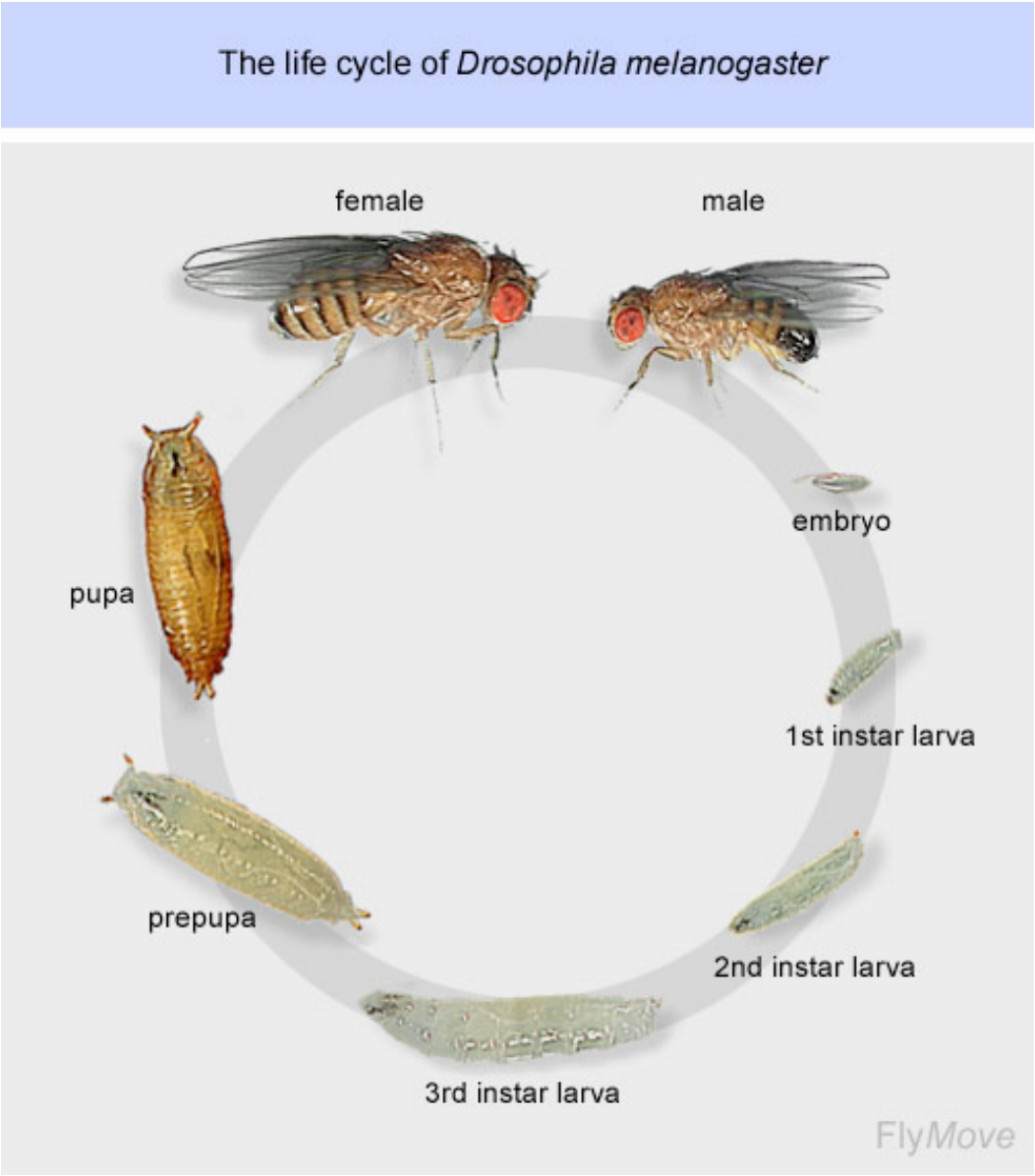


Plate 1.10: Life cycle of Drosophila melanogaster (courtesy : FlyMove)

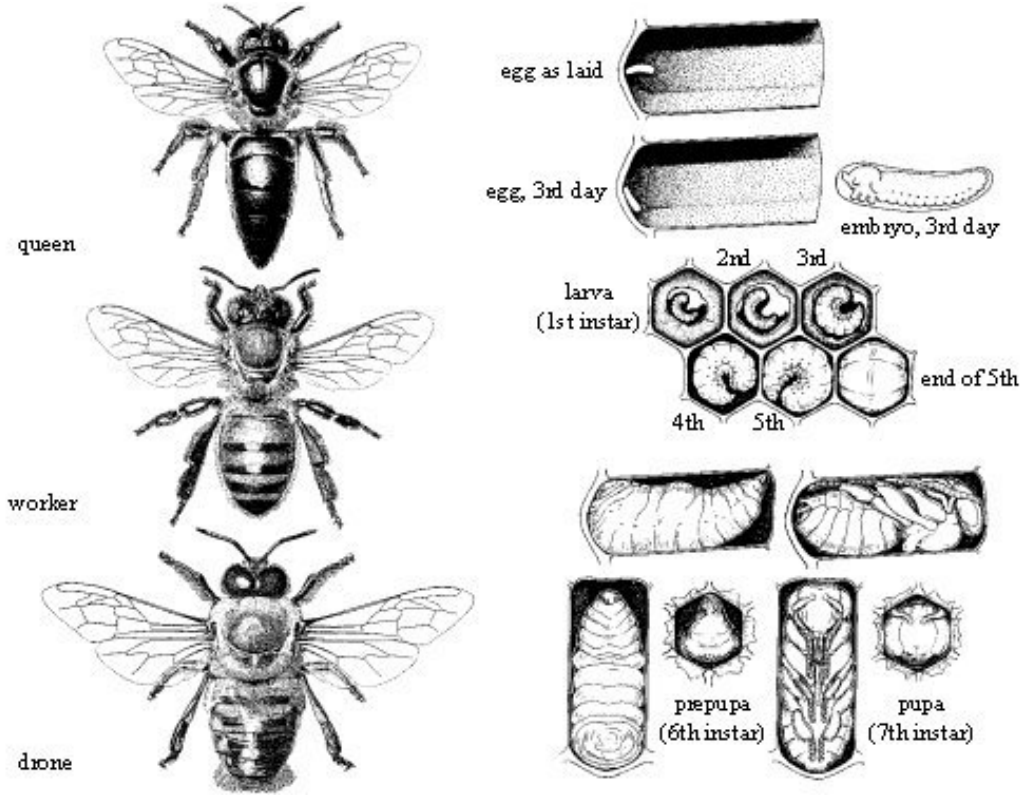


Plate 1.11: Life cycle of *Apis mellifera* (courtesy : PhD thesis of Denholm)

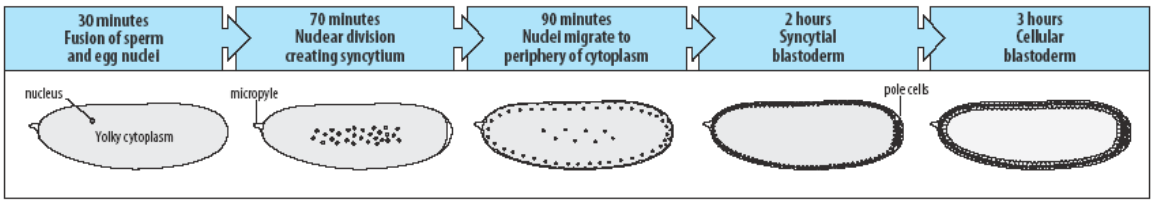


Plate 1.12: Cleavage in early *Drosophila melanogaster* embryo and formation of blastoderm. (Courtesy: Gilbert *et. al*, Development of the *Drosophila* body plan)

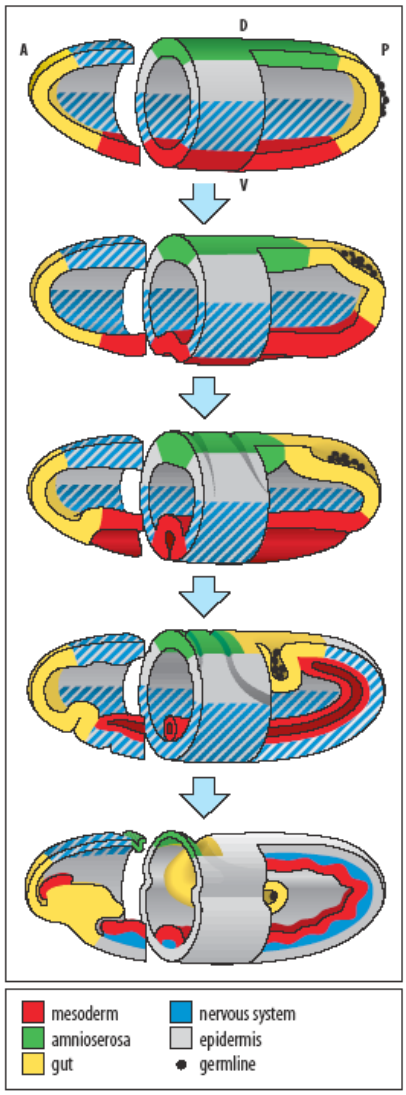


Plate 1.13: Gastrulation in *Drosophila melanogaster.*

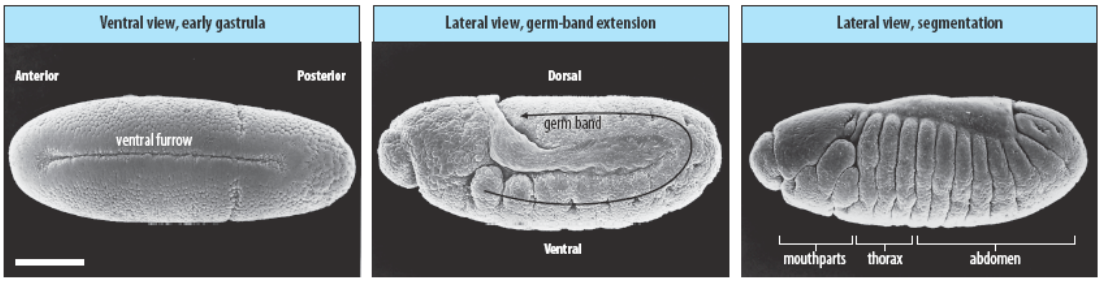


Plate 1.14: Gastrulation, germ band extension and segmentation in *Drosophila melanogaster* embryo.

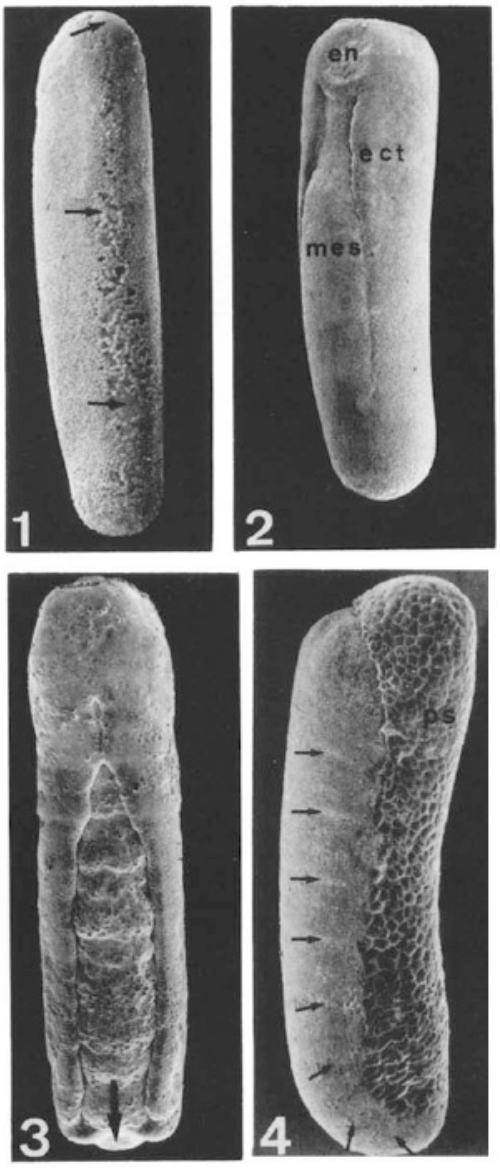


Plate 1.15: Blastoderm, gastulation and early germ band stage of *Apis mellifera.* (adapted from Fleig and Sander, 1986)

1. Late blastoderm stage. Dorsal strip is shown with arrows
2. Ventro lateral view of the gastrulation stage. Prospective mesoderm, endoderm and ectoderm have been indicated
3. Advanced gastulation stage
4. Early germ band stage. Prominence of every second segmental groove has been indicated with arrows.

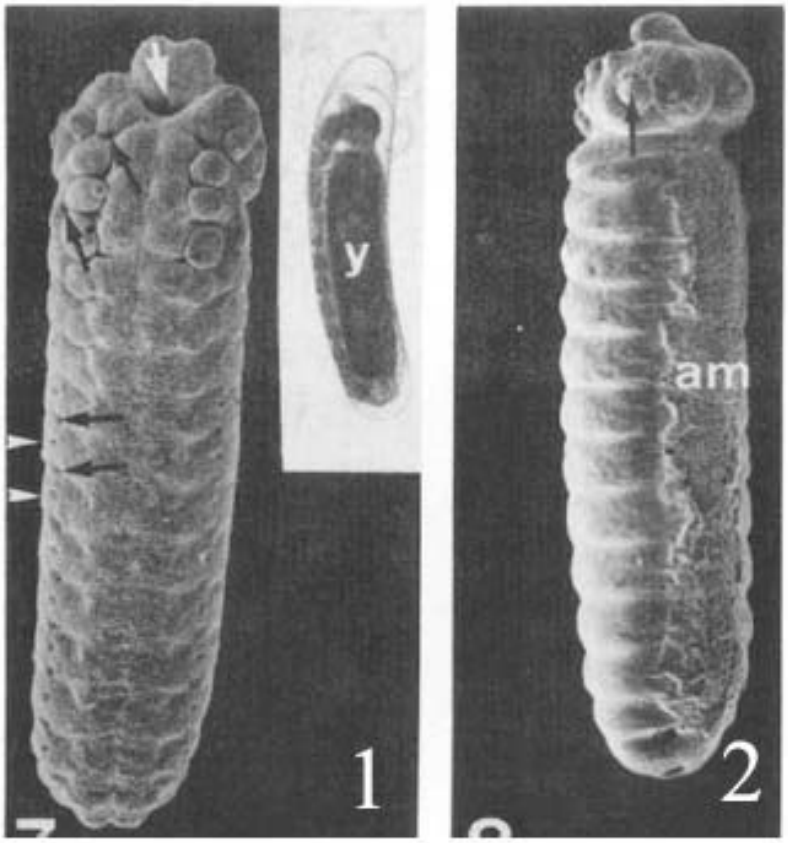
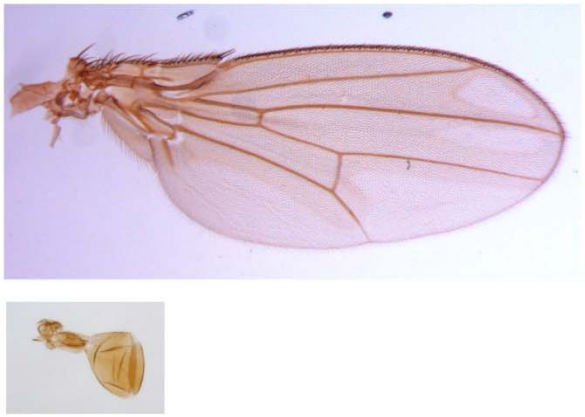
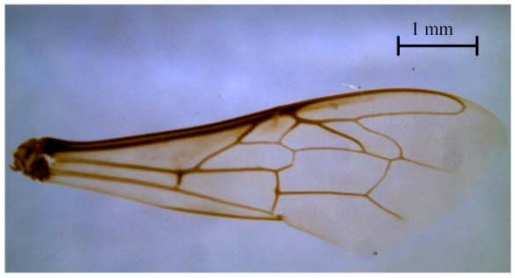


Plate 1.16: Advanced germ band stage and dorsal closure. (Courtesy: Fleig and Sander , 1986)

1. Advanced germ band stage (white arrows on labrum mark the invaginations giving rise to stomatogastric nervous system; black arrows in the head show pits of tentorium; black arrows on abdomen mark the invaginations related to ocnocyte formation while white arrows in the abdominal region mark the tracheal pits).
2. Dorso lateral view of embryo during dorsal closure



Wing & haltere of fruit fly



Fore & hind wing of honey bee

Plate 1.17: Dorsal second and third thoracic appendages of *Drosophila melanogaster* and *Apis mellifera*

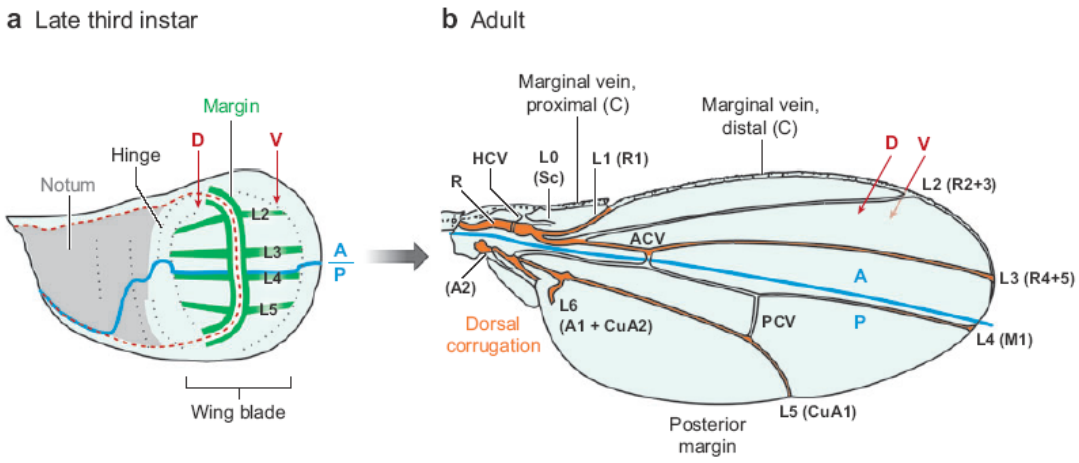


Plate 1.18: Late third instar wing disc and adult wing blade (image Courtesy: Blair, 2007).

a) Position of longitudinal veins in late third instar wing imaginal disc.

b) Veins in adult wing

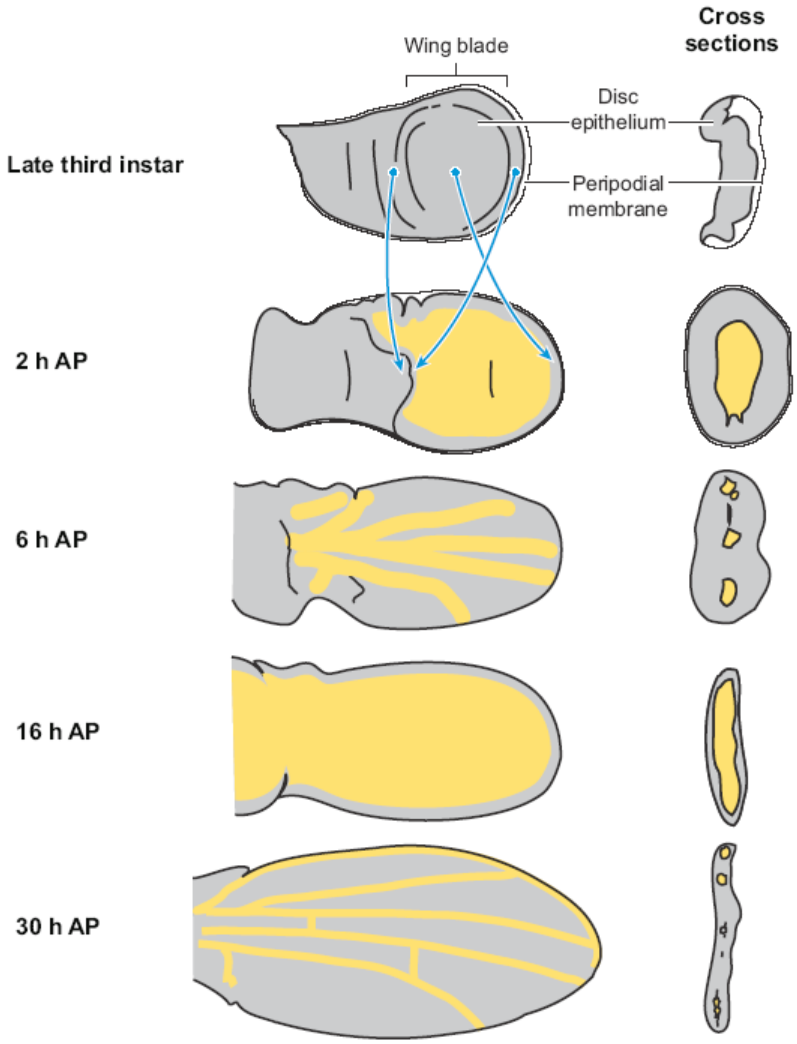


Plate 1.19: Morphogenesis of the wing and the formation of veins and inter veins (courtesy: Blair, 2007)

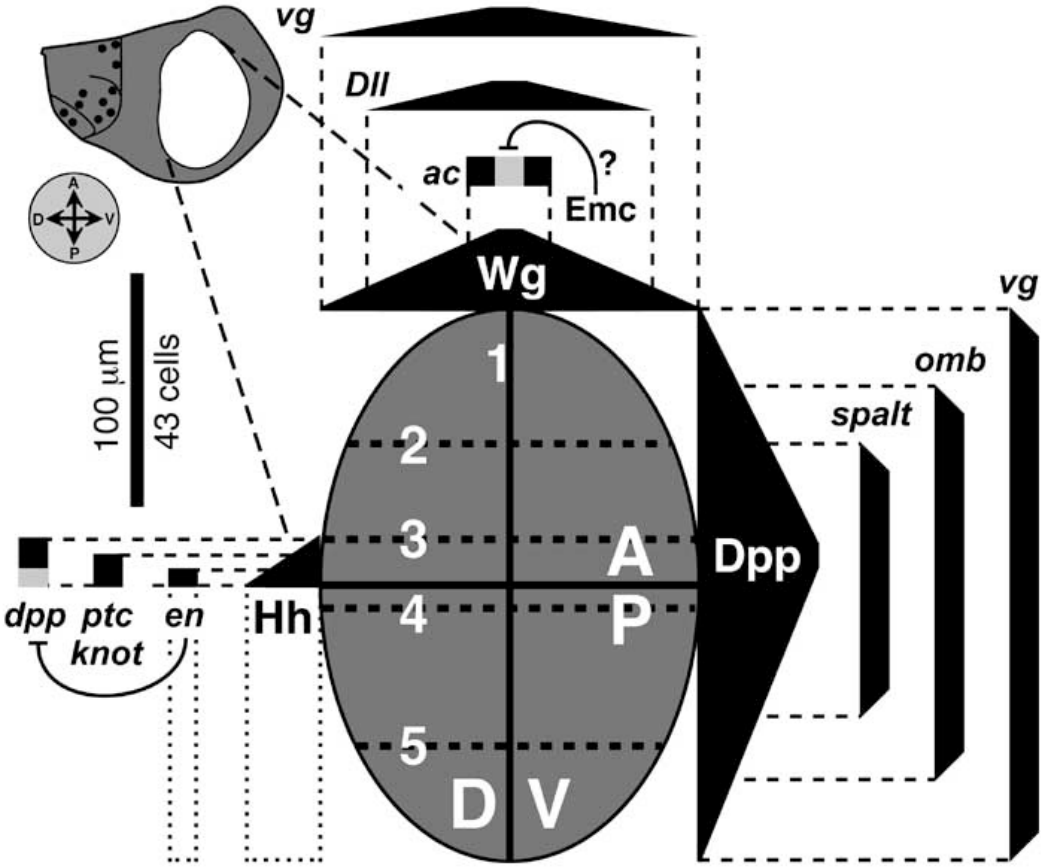


Plate 1.20: Activity of morphogens along the A-P and D-V axis of Drosophila wing imaginal disc (Courtesy: Lewis Held).

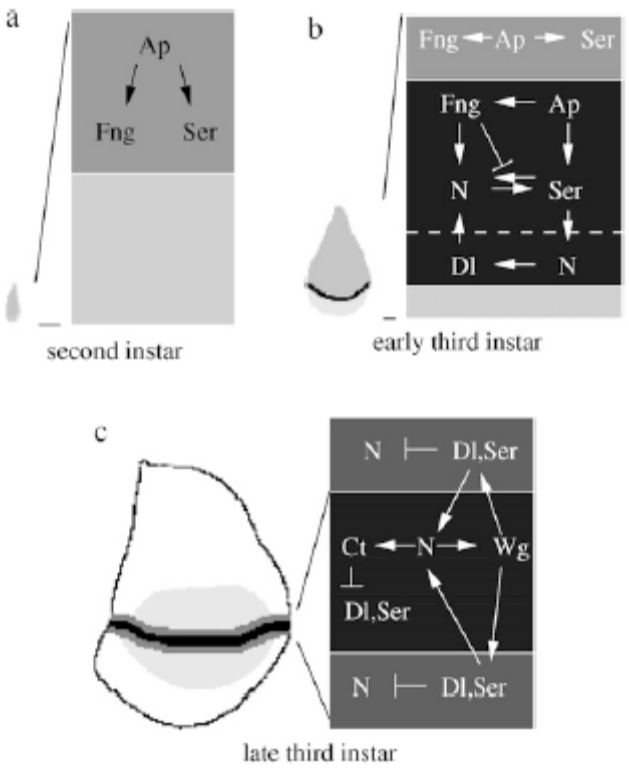


Plate 1.21: Regulatory interactions defining the D-V boundary (Courtesy: Gonzalez *et. al.*, 2006).

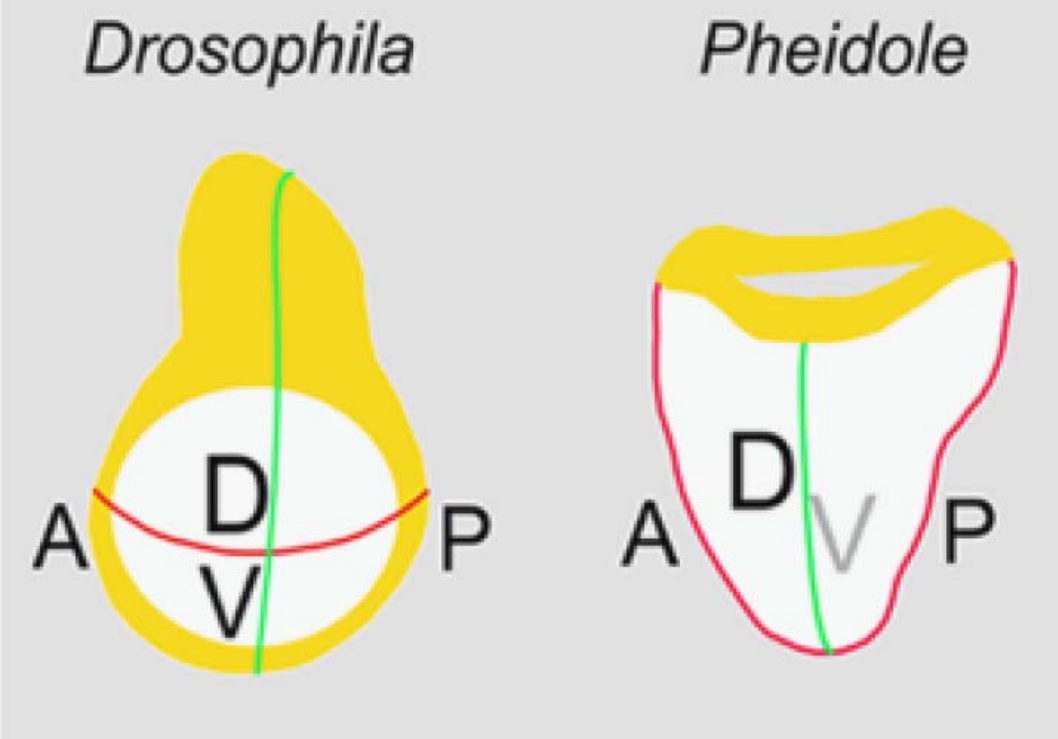


Plate 1.22: Schematic diagram of *Drosophila* and *Pheidole* wing discs. Pouch is shown in white while hinge is shown in yellow. The red line shows the D-V boundary while the green line shows the A-P boundary.

Courtesy: Shbailat *et.al.*, 2010

Chapter2

Chromatin Immunoprecipitation & Sequencing

Introduction

In order to identify direct targets of Ubx in vivo in the hind wing discs of *Apis mellifera*, a technique called chromatin Immunoprecipitation (ChIP) was used. The binding of transcription factors to various regions of chromatin which may be in different dynamic states at various stages, plays a pivotal role in the biology of organisms. In order to understand various biological processes that are controlled by a particular transcription factor at a global level, it is essential to have a genome wide understanding of the protein –DNA interactions. The main tool for understanding these interactions in vivo is ChIP (Solomon *et. al.*, 1988). ChIP has been coupled with conventional PCR or qPCR to determine if a particular region has been bound by a transcription factor or not. Cloning of DNA fragments immunoprecipitated by ChIP also has been used as a technique to identify novel binding sites of transcription factor. The ability to sequence millions of small DNA fragments at a low cost with a high efficiency with the help of modern techniques has made it possible now to sequence almost all the DNA fragments enriched during ChIP. The technique, known as ChIP-seq, helps in a non biased study of the binding sites at a global level which was not possible earlier.

2.1: Chromatin Immunoprecipitation

In a ChIP experiment, the DNA binding protein is first cross linked to chromatin using formaldehyde and the chromatin is sheared by sonication into short fragments about 200- 500 bp in length. A specific antibody against the protein in question is then used to immunoprecipitate the DNA- protein complexes. The DNA and protein cross links are reversed and the DNA is subsequently purified. The resulting DNA is enriched for the regions that are bound by the protein/ transcription factor. In few cases, the protein is not cross linked to the DNA by formaldehyde (and hence the fragmentation is achieved by treating the chromatin by MNase); the process is called

N- ChIP or native-ChIP. The widely used method where the DNA is cross linked to the protein using formaldehyde is called X- ChIP.

Plate 2.1:An overview of ChIP protocol

2.2: Arrays for detection of enriched DNA during ChIP

The resulting DNA fragments can then be hybridised on a tiling array or can be sequenced on various platforms. A tiling array contains almost the entire genome of an organism on small chips. Small probes (around 58 -60 bp in length) mapping the entire genome are either synthesized directed on the surface of array by various methods or printed on the surface of the array. Various types of tiling arrays are available depending on the study that needs to be done. Two types of unbiased whole genome tiling arrays are available- one contains partially overlapping or non-over lapping probes to cover the entire genome while the second contains non – overlapping probes spaced at regular intervals to interrogate the entire genome.

Biased whole genome array designs include expression arrays, splice junction arrays and exon scanning arrays that contain only nucleotide probes for the predicted or the known features of the genome.

A tiling re-sequencing array represents each nucleotide of the reference genomic sequence with a set of eight oligonucleotide probes (**plate 2.2**).

Any of the two non-biased tiling arrays can be used to hybridise the chip isolated DNA and subsequently map the DNA fragments to the genome. High reproducibility amongst arrays, unbiased and complete genome coverage and multiple potential overlapping probes representing the transcription factor binding sites are the advantages of using tiling arrays. Theoretically, a resolution of 25 bp can be achieved using tiling arrays but the high costs associated with achieving such a resolution makes it a non practical approach. (Reviewed in Mockler *et. al.*, 2005).

2.3: Sequencing of the chipped DNA fragments

Various platforms are now available, which can sequence the chipped DNA and generate short reads, which can be mapped to the genome. The greatest advantage of sequencing is the resolution that can be achieved with the technique. Arrays too can be tiled to achieve greater resolution but the density at which they must be tiled to achieve the desired resolution raises the cost several fold- especially for large genomes. Hybridisation step for the use of microarray generates a lot of noise, which is eliminated in sequencing. The genome coverage of a tiling array is always limited by the sequence that is available on the array while the same is never a limitation for sequencing.

The sequencing technology too has some unwanted artefacts- especially in the sequencing reads towards the end. This can, however be eliminated by using better bioinformatics approaches. Various different platforms are now available for sequencing of DNA-the most popular ones being

a) 454 Genome Sequencer FLX instrument (Roche Applied Sciences)
b) Illumina (Solexa) genome analyzer
c) Applied Biosystems (ABI SOLiD system)

454 Genome Sequencer FLX instrument: The instrument is based on the principle of pyrophosphate detection, which was described in 1985 (Nyren e*t. al.,* 1985) and the principle was used for sequencing DNA in 1988 (Hyman, 1988).In this system, a library is prepared by ligation of specific primers/adaptors to the DNA fragments to be sequenced. These adaptors are then used to bind the fragments to a bead where an emulsion PCR is carried out. The PCR reaction is carried out in water droplets that contain the bead with the DNA with the PCR reagents contained in an oil droplet within the water drop.

Once the PCR is complete, the library is placed on an optical fibre chip. Each bead with hundreds of amplified fragments is placed in a glass fibre whose position in the

chip is well known. Now, only one unlabelled nucleotide, together with other PCR components is added to the bead so that synthesis may take place. The release of pyrophosphate group is detected as a light signal. Knowing the identity of the nucleotide supplied, one can interpret the sequence based on the signal emitted. The system can achieve a read length of around 400 base pairs. (**Plate 2.3**)

The Illumina (Solexa) Genome Analyzer: The instrument is based on sequencing – by synthesis chemistry. Library is prepared by ligating the DNA fragments on both ends to adaptors.

The library is then immobilised (after denaturation of DNA) at one end of a solid support, which is coated with adaptors and complementary adaptors. A bridge of single stranded DNA is created as the adaptors on both end of DNA hybridise to their complementary adaptors on the surface. Amplification is then carried out to generate random clusters of about 1000 single stranded DNA fragments in each cluster. The reaction mixture for sequencing reaction containing four reversible terminator nucleotides each labelled with a different dye is then applied to the support. After incorporation into the DNA, the position of the dye is detected using a CCD camera and then the terminator group at the 3' end of the dye is removed which is followed by a new synthesis cycle. The sequence read length obtained by this cycler is around 35 bp (in recent times it has gone up to 100 bp).**Plate 2.4** shows the workflow used by Illumina Sequencer

The Applied Biosystems (ABI SOLiD system): The system uses chemistry based on ligation for sequencing the DNA. The DNA fragments are first ligated to adaptors and then attached to a bead. Emulsion PCR is then carried out in a water droplet, which is contained in oil emulsion.

Next, the amplified DNA is denatured and one bead is loaded on a glass support. First a primer is hybridised to the adaptor. Now, oligonucleotide octamers are taken, the doublet of fourth and fifth bases is characterised by any one of the four labels present at the end of octamer. These octamers are then hybridised to the DNA fragments and

ligation mixture is added. After the detection of signal, the identity of fourth and fifth base is determined. The ligated octamer is now cleaved off after the 5^{th} base and now a new cycle is repeated, which gives the identity of 9^{th} and 10^{th}bp. The sequencing reaction is continued in this way using primers shorter than the earlier by 1 bp, thus determining the sequence at 3 and 4^{th}position, 8^{th} and 9^{th} position and so on. The length of sequence reads obtained earlier was around 35 base pairs but has gone up subsequently.**Plate 2.5** shows an overview of the reactions used by ABI SOLiD

For the purpose of sequencing in this study, I used Illumina sequencing platform.

2.4: Requirements for Chromatin Immunoprecipitation experiment

As outlined earlier, ChIP essentially involves cross-linking DNA to proteins in vivo and immunoprecipitating the DNA protein complex using antibodies specific to the protein in question. Two important criteria therefore come into the pictures, which need special attention- antibody and the tissue type to be studied.

Antibody: Chromatin immunoprecipitation requires that the antibody must recognise the protein epitopes in solution and even when the protein is bound to DNA. Binding of protein to DNA may mask few epitopes and also change the configuration of the protein slightly so that very few of the epitopes present in the native protein may actually be available for the antibody. In order to recognise the protein in most configurations and in solution, it is essential that polyclonal antibodies should be used. As also described earlier, that all the Hox proteins contain a common homeodomain and YPWM motif, it becomes essential that a recombinant protein is used for generation of antibodies that doesn't contain these domains- producing a protein that would be specifically recognise the protein in many configurations.

It is also important that the sera generated be purified for subsequent applications for ChIP to avoid variations between replicates. The antibody in the study was produced by immunising rabbit against the antigen. Hence, protein A column was used for purification as rabbit IgG binds strongly to protein A. Protein A is a cell wall protein

from *Staphylococcus aureus* with a molecular weight between 35- 50 kDa. Protein A binds to IgG at higher pH (pH 9.0). At this pH most other components of the sera donot bind to protein A and hence can be washed off. IgG can be subsequently washed from protein A by lowering the pH to around 3. At this pH however, IgG is highly unstable and hence the pH is immediately neutralised after elution.

Tissue: The most important criteria after selection/ generation of antibody is to select the tissue type and stage in order to address the question. As already stated, Ubx serves to modify the fate of the third thoracic wings in insects to various extents. The modification is essentially a process of differentiation, which beings in *Drosophila* during late third instar larva. During this stage, in the wings of fruit fly a transcription factor called Cut expresses along the dorso-ventral boundary. The protein is a very specific marker in *Drosophila* indicative of the beginning of differentiation.

In order to compare two different insect species for Ubx targets, it is essential that we compare the tissues isolated at comparable developmental stages. This stage should be the one where differentiation process has just begun. As this stage is the late third instar of larvae in fruit fly and the stage in wing imaginal disc is characterised by the appearance of Cut, the same protein was used as a marker to identify the stage at which ChIP should be carried out in honey bees.

Wing patterning network seems to have existed for more than 300 million years (Abouheif *et. al.*, 2002) and wing polyphenism seems to have arose around 125 million years ago. As the wing patterning network seems to have originated before the origin of hymenopterans, many of the wing patterning genes are conserved between dipterans and hymenopterans. Wing polyphenism, provides a useful way of recognising the wing buds of hymenopterans- in different castes of few hymenopterans there is either reduced or no growth of wings. When traced back to the larval stage, the dorsal thoracic buds/ discs in these castes are either rudimentary or do not appear at all while in the winged castes the same discs/ buds are bigger in size.

Type of honey bees that can be used for the study

As already stated in the previous chapter that the European Honeybee whose genome has been sequenced is an ideal model for the study. An important point that needs to be taken care of is the diversity amongst the population of honeybees that are used for chromatin immunoprecipitation studies. As a single queen lays eggs in one hive, therefore in order to minimise genetic variation, one can use the larvae from a single hive which houses a queen that has been artificially inseminated with semen from a single drone.

In order to carry out bioinformatics analysis, after sequencing of the chipped DNA, it is essential that the sequenced chipped DNA fragments be aligned to the genome. However, due to the complexity of the bioinformatics analysis, it is always useful to have as less variation as possible within a particular experimental sample. The genome of the honey bee, *Apis mellifera* which has been sequenced belongs to strain DH4 and was obtained from an apiary in United States. The advantage of using honeybee is the fact that a single queen lays eggs in the entire hive and hence all the progeny inherit similar genotypes. The queen raised in wild type environment, however is mated by multiple drones during nuptial flights and hence the eggs she lays, have different genotypes. In order to have honey bees of very similar genotype, queens that were artificially inseminated using sperms of a single drone were used. This minimised the chances of having variability within the population of honey bee that were used for ChIP experiments. For a single ChIP experiment, larvae from one hive only (which were offspring of a single queen) were chosen. These hives were maintained at the Department of Genetics, Heinrich Heine University, Dusseldorf, Germany.

Using polyclonal antibodies specific to Ubx of honey bees ChIP was carried out on the hind wing buds as described in the following section.

Materials & Methods

2.5: Honeybees used for the study

For initial identification of honey bee wing discs and for antibody staining, the bees were collected from various bee keepers in Karnataka. Later, bee hives were maintained at Pune. The bees were fed with sugar solution and pollen substitute, while they were maintained in Pune.

All the larvae that were used in the Chromatin immunoprecipitation studies were obtained from one single hive housing a colony that were offspring of a single queen artificially inseminated using semen from a single drone.

2.6: Identification of various larval stages of honeybee

The larvae of honeybees were staged according to the morphological features (described in the first chapter & adapted from the thesis of Stephen Eichmuller). Essentially, the larvae were identified or staged to a particular instar based on their size and curvature in the tiny hexagonal cells.

2.7: Identification of wing buds/ discs in the larvae of honeybee

Wing development network has been conserved in various insects across million of years. In order to precisely identify the wing disc in honeybee, a comparison of wing discs in other hymenopterans was made. Also, few light microscopy studies are available, which show the rough location of honeybee wing disc. Comparing the shape and using the dorsal location in the second and third thoracic segments, the wing buds were identified.

2.8: Antibody staining

Antibody staining was done according to the protocol described in Patel *et.al.*, 1989 with few modifications. The larvae of honeybee were dissected and fat bodies

removed to expose the wing imaginal discs. The larval heads with wing discs were fixed in 4% formaldehyde for 25 minutes at room temperature. They were subsequently washed thrice with phosphate buffered saline, pH 7.4 containing 0.1% Triton 100-X (PBTX -0.1), each wash lasting for 10 minutes. The heads were then rewashed with phosphate buffered saline, pH 7.4 (PBS) for 10 minutes and now the wing discs were dissected under a light microscope. The wing discs were blocked using 1 % bovine serum albumin (BSA) in PBTX-0.1 for 30 minutes at room temperature. The antibody at desired concentration in PBTX -0.1 (containing 0.5 % BSA) was incubated with wing discs overnight at 4 C with slight shaking. The wing discs were then washed thrice with PBTX-0.1 at room temperature, each wash lasting minimum 10 minutes. The discs were then re-blocked using 0.5 % BSA in PBTX-0.1 for 30 minutes at room temperature. Secondary antibody in PBTX-0.1 containing 0.5 % BSA was incubated with the wing discs for 1 hour at room temperature and then the wing discs were washed again using PBTX-0.1. The wing discs were incubated with DAPI for 20 minutes, stored in PBS and mounted in 50% glycerol in PBS.

In order to detect the antibody by peroxidise reaction; DAB kit from Vector Labs was used. Staining with primary antibody and washing was carried out as described earlier. However, instead of fluorescent secondary antibody, secondary antibody conjugated with HRP was used and detection was carried out using the solutions of the kit according to manufacturer's guidelines.

Antibodies used in this study:

Anti Cut (DSHB) : The antibody was used at a dilution of 1:5

Anti Extradenticle (DSHB): The antibody was used at a dilution of 1:5

Anti Spalt: The Antibody was used at a dilution of 1:50.

Anti Engrailed (DSHB): The antibody was used at a dilution of 1:5

Anti Ubx (this study): The antibody was used at a dilution of 1:3000.

2.9:Generation of polyclonal antibodies against recombinant Ubx protein of *Apis mellifera*

In order to generate polyclonal antibodies against N- terminal region of Ubx of *Apis*, first Ubx cDNA clone was isolated by RT-PCR, the recombinant clone bearing *Apis* Ubx fragment (Ubx excluding the homeodoamin and YPWM motif) was generated in pET 15b vector. The resultant recombinant protein had a N- terminal poly-His tag which enabled easy purification of the protein using Ni-NTA matrix.

Step1: Total RNA isolation from honeybee larvae

In order to isolate RNA, around 25mg of *Apis mellifera* larval tissue was crushed in liquid Nitrogen and mixed with 1mL of TRIzol. The tissue was properly homogenised in TRIzol and the insoluble material was removed by centrifugation. The homogenised sample was incubated at room temperature for 5 minutes to permit complete dissociation of nucleoprotein complexes. 0.2 mL of chloroform was added to the mixture, shook vigoursly, incubated at room temperature for 10 minutes and then centrifuged to separate the phases. The upper aqueous phase was removed after centrifugation and RNA was precipitated by adding 0.5 mL isoprapanol to it. The RNA precipitate was collected by centrifugation, washed with 70% ethanol and then dissolved in DEPC- treated water. The integrity of RNA was assessed by running an aliquot on an agarose gel while the purity was assessed by measuring its absorbance at 260, 280 and 230 nm.

Step 2: Preparation of cDNA

Reverse Transcription was carried out using MMLV Reverse Transcriptase (Promega). Two micrograms of RNA was mixed with 0.5 µg of anchored dT primer and the mix was heated at 70° C for 5 minutes to denature the secondary structure of RNA. 5 µL of 5X Reverse Transcription buffer and 1.25 µL of dNTP was added and the mix was incubated at 25° C for 10 minutes. 1.5 µL of Reverse Transcriptase was added and incubated at 42° C for two hours. The reaction was then stopped by

heating at 70 °C for 15 minutes and stored at 4°C. The cDNA was then used as a template to amplify the N – terminal region of *Apis* Ubx.

Step3: Amplification and ligation of N- terminal region of Ubx

In order to get antibody specific only to Ubx protein, the homeodomain and the YPWM motif (which are common amongst different Hox protein) was removed. Primers specific to the N- terminal fragment of Ubx (excluding Homeodomain and YPWM motif) were designed. The forward primer had a recognition site for Nde1 and the reverse primer had a recognition site for Bam H1 for easy directional cloning into pET 15b.

The following mix was used for PCR:

MilliQ	:	14.2 µL
10 X Pfu Buffer	:	2.5 µL
$MgCl_2$	:	2.5 µL (final conc of Mg :2.5mM)
Forward Primer	:	1.25 µL
Reverse Primer	:	1.25 µL
dNTP	:	0.8 µL
cDNA	:	1.0 µL
DMSO	:	1.0 µL
Pfu (Stratagene)	:	0.5 µL
Total volume	:	25 µL

The following program was used for the PCR

95 °C for 5 minutes

30 cycles of

95°C for 30 seconds

58°C for 30 seconds

68°C for 2 minutes

Final extension of 5 minutes at 72°C

Expected product size: 560bp

The amplified product was electrophoresed and the band was eluted and purified using Gel purification kit (QIAGEN) according to the manufacturers' protocol.

In order to clone the amplicon in pET 15b, both the amplicon and the vector were digested using Bam H1 and Nde1 (New England Biolabs) according to the conditions described by the manufacturer. The digested product was electrophoresed on anagarose gel and eluted using Gel extraction kit (QIAGEN). A ligation reaction with insert to vector ratio of 3:1 was set up using T4 DNA ligase (Promega) at 16°C overnight. The following formula was used to calculate the amount of insert needed for a given reaction for a fixed amount of vector (according to the vector: insert ratio):

$$\text{ng of insert} = \frac{\text{ng(vector)} \times \text{kb (insert)} \times \text{(Insert:Vector molar ratio)}}{\text{kb(Vector)}}$$

The following ligation mix was used for the reaction:

10 X ligase buffer	:	1.6 µL
Insert	:	10.5 µL

Vector	:	3.0 μL
Milli Q	:	0.4 μL
T4 DNA ligase	:	0.5 μL
Total		16 μL

The ligated mix was used to transform DH5 alpha cells (Sambrook *et. al*, 2001). The resultant colonies were screened by colony PCR for positive clones bearing the insert. Few positive colonies were picked, their plasmids were isolated and sequenced. The correct clone was used in further studies.

Step 4: Expression and purification of recombinant protein

The clone was then used to transform BL21DE3 cells. A single isolated colony was used to inoculate terrific broth containing (100μg/mL ampicillin).The culture was grown at 37°C in a water bath at 150 rpm overnight. This primary culture was used to inoculate 200 mL terrific broth containing 100μg/mL ampicillin. The culture was grown at 37°C in a water bath and was induced with 1mM IPTG (final concentration) in the log phase of growth (O.D ~0.65-0.9). The cells were grown for an additional period of 4 hours after induction at 37°C. The cells were then pelleted by centrifuging them at 4000 rpm for 10 minutes. The cell pellet was mixed with 6.5 mL denaturing lysis buffer (30 X).

The lysate was then kept for shaking on an orbital shaker at 29°C for 2 hours. It was then centrifuged at 14000 g for 30 minutes to settle all the insoluble particles. The supernatant was applied to Ni- NTA matrix, which had been prewashed and equilibrated with lysis buffer. 4 mL lysate was applied to 1mL Ni-NTA matrix. The lysate matrix mix was rotated at room temperature for 2 hours. The mix was then packed into a column. The column was washed with wash buffer , pH 6.3 five times.

The column was then washed with wash buffer, pH 5.9 five times. Finally the protein was eluted with elution buffer, pH 4.5 .

Step5: Generation of polyclonal antibodies against the recombinant protein

The purified protein was dialysed against water (supposed to increase the antigenicity of the protein) and then an aliquot was electrophoresed on a SDS gel to establish the integrity of the protein after dialysis against water. A rabbit was selected for antibody generation, from which around 3 mL of blood was removed as pre immune sera. The dialysed protein was then mixed with freund's complete adjuvant and it was injected intradermally to the rabbit (day 0). First booster was administered to the rabbit on day 34. Protein was mixed with equal volume of freund's incomplete adjuvant and injected into the rabbit. Second booster was administered in a way identical to the first booster on day 50. A test bleed was removed on day 57 and checked by Western blot hybridization. Once the antibody was confirmed, a first bleed (around 20 mL) serum was removed from the animal. Third booster was given on day 86. Second bleed was removed on day 94. The antibodies, thus obtained, were used for subsequent applications.

Step 6: Testing the specificity of antibody

In order to establish the specificity of the antibody – a Western blot hybridization was performed with the antibody against the whole larval lysate of honey bee. Embryos of honeybee were homogenized in cell lysis buffer and then were mixed with 2X gel loading dye followed by boiling for about 12 minutes to rupture the cells completely. A 10% Polyacrylamide gel was casted and electrophoresis was carried out as per the standard procedure. Following electrophoresis, the proteins were transferred on a PVDF membrane at constant current of 300mA for 2 hours at 4 °C. In order to carry out the transfer, the gel and pre-wet PVDF membrane was first equilibrated in transfer buffer for 30 min at room temperature. The gel and membrane were then assembled in a sandwich and the protein was transferred in a submerged transfer apparatus at 300mA for 2 hours at 4^{o}C. The membrane was then removed and washed

with TBS four times. The membrane was then blocked with 4% milk in TBST O/N at 4 ^{o}C. Primary antibody (anti-Ubx) was added to the blot at 1:10,000 dilution and kept at room temperature on shaker for 2 hours. The blot was then washed with TBST for 20 minutes. The blot was then blocked with TBST plus 2 % BSA for 30 minutes. Secondary antibody was added TBST containing 2 % BSA (1:5000;Zymed) and incubated for 2 hours at room temperature.

The blot was washed with TBST and visualised on chemiluminescent imager.

2.10: Chromatin Immunoprecipitation

The whole protocol has been subdivided into smaller sections

Step 1: Chromatin preparation

Larvae were collected from a single hive with artificially inseminated queen. All the larvae that were used were of early fifth instar. The larvae were dissected and cut open in ice cold PBS. The cut larvae were incubated in 1.5 % formaldehyde for 10 minutes at room temperature and were then washed thoroughly with PBS. Fat bodies were dissected out and the wing discs were exposed. The larvae with exposed wing discs were then fixed in 1.5% formaldehyde for 30 minutes at room temperature. The reaction was quenched by adding one tenth the volume of 1.375M glycine in PBS and shaking it for 10 minutes. The larval heads were washed thrice in ice cold PBS to get rid of formaldehyde. The wing buds were dissected out. For each reaction, a wet tissue volume of around 30 µl was used (90 fore wing discs and/ or 130 hind wing discs). The discs were centrifuged at 1000 rpm at 4º C to allow the discs to settle down. Around 300 µl of swelling buffer was added to the discs (10 times the volume of discs). The discs were incubated on ice for 15 minutes and were dounced intermittently to rupture the cell membrane. The ruptured cells were spun at 2000 rpm for 5 min at 4ºC and pellet re-suspended in 300 µl of sonication buffer. The mix was dounced twice and then sonicated in a diagenode water bath with the following settings

30 s ON

30 s OFF

Total time: 5 + 5 minutes

Power: HIGH

The sample was centrifuged at 14000 rpm for 15 minutes at 4 ^{o}C. Supernatant was removed and this supernatant contains chromatin. It was flash frozen using liquid nitrogen and stored in -70^{o}C.

Step2: Immunoprecipitation

Preparation of Protein A sepharose

0.25 g of Sepharose was weighed (for a final volume of 1mL media) and suspended in distilled water by gentle swirling. The mix was allowed to settle and water was decanted out. To achieve rapid settling the mix was spun at 500 rpm. This process was repeated five times so that the media was washed with a total of 100mL water. A slurry was prepared in binding buffer containing 75 % media, 25 % buffer (in this case dilution buffer)

Purification of antisera

Protein A column from Montage was used to purify the anti Ubx and pre immune IgG sera. The purification was carried out according to the manufacturer's instructions. Briefly, PROSEP-A media was equilibrated with 10 mL binding buffer by centrifuging the spin column at 500g for 5 minutes. The sera was then pre-cleared by filtering through a 0.22 μm filter. The filtered sample was then diluted 1:1 (v/v) in binding buffer. The sample/ sera was loaded on the spin column and centrifuged at 150 g for 20 minutes at 4^{o}C. The spin column was then washed with 20 mL of binding buffer by centrifuging the spin column at 500 g for 5 minutes at 4^{o}C. The bound IgG was now eluted with 10 mL elution buffer into a fresh centrifuge tube

containing 1.3 mL neutralisation buffer to bring the sample to neutral pH by centrifuging the column for 5 minutes at 500g at 4^{o}C.

The IgG fraction was concentrated using Amicon Ultra 15 centrifugal device with 30000 NMWL. The concentrated antibodies were stored by stabilising them in 50% glycerol at -70^{o}C. The integrity of the antibody was assayed by running a small aliquot on SDS gel. Both the heavy and light chains were clearly visible on the gel. The antibodies were quantified by measuring their absorbance at 280 nm. The concentration was achieved by multiplying the absorbance by 0.72 to get the concentration in mg/ml.

Processing for ChIP

Protein A Sepharose was washed with sonication buffer thrice. Sepharose was incubated for four hours in sonication buffer containing 1 mg/mL BSA. Chromatin, (earlier stored in – 70 ^{o}C) was thawed on ice and divided into 100µl aliquots for immune precipitation. The final volume of each aliquot was made to 1 mL using dilution buffer. BSA was added to make a final concentration of 1 mg/ mL. The lysate was pre cleared by incubating it with protein A sepharose for two hours in the cold room with gentle shaking. The samples were centrifuged at 3800 g for 5 minutes at 4 ^{o}C. Supernatant was taken – this supernatant contains pre cleared chromatin. One aliquot was used for pre immne IgG and two others were used for anti Ubx IgG pull down. Three µl of antibody was added to all three aliquots and rotated for 2 hours in cold room. 40 µl of protein A sepharose was then added and incubated overnight with constant rotation at 4^{o}C. The mix was centrifuged at 3800 g for 5 minutes at 4^{o}C. The beads were then washed 2x each with 1 mL sonication buffer, 1 mL wash buffer A, 1 mL wash buffer B and 1mL TE. 100 µl elution buffer was added to each of the three aliquots and placed on a vortex mixer for 10 minutes. The tube was centrifuged and the supernatant was removed and stored. Another round of elution was carried out with 75 µl of elution buffer. Both the fractions were mixed and saved.

(Washing was done in all steps by incubating the tubes for 10 minutes in cold room on rocker and then centrifuging the tubes)

Step3: Reverse Crosslinking to check chromatin size and for isolation of DNA

The volume of eluate from each aliquot was made to 400 µl. To it 21µl of 4 M NaCl was added and incubated at 65 ^{o}C for 5 hours. 1 µl of RNaseA (10mg/mL) was added and incubated at 37^{o}C for 1 hour. 4 µl of 0.5 M EDTA and 2µl of proteinase K (10 mg/mL) was added and incubated at 42^{o}C for 2 hours. DNA was then extracted using phenol: chloroform :iso amyl (25:24:1) once followed by washing using chloroform : iso amyl alcohol (24:1) once. One µl of glycogen (stock 20mg/ml) and 1/20th the volume 4M NaCl (final conc 0.2 M) and 2.5 volumes ice cold ethanol was added. DNA was allowed to precipitate overnight at – 20^{o}C.It was centrifuged, washed and resuspended in 10mM Tris,pH 8.

Results & Discussion

2.11:Sample collection

Honey bees were collected from various locations in Karnataka and later maintained at IISER, Pune. (**Plate 2.6** shows various locations from where the bees were collected while **plate 2.7** shows the organisation of hive maintained locally). A good population of bees was present in areas where there was abundant supply of pollen and nectar was present. Such locations in India are restricted to fields where floral plants like Sunflower, Mustard or Rubber are grown. Hence for collection of bees for initial studies, such as determination of developmental stages, immunohistochemistry etc, the samples were collected from various bee-keepers who moved according to vegetation and season. While maintaining the honeybees in Pune, artificial food supplement was given to honeybees in the form of sugar solution and ground, roasted pulses mixed with honey.

2.12: Identification of wing discs/ wing buds of *Apis mellifera*

The wing buds were located on the dorsal side of the larvae with the leg discs lying on the ventral side (**plate 2.9**). The wing buds were then stained using DSHB antibodies that have been shown to cross react with other insect proteins. Some of the genes whose expression patterns were tested are *cut* (**plate 2.10**), *spalt* (**plate 2.11**), *engrailed* (**plate 2.12**) and *extradenticle* (**plate 2.13**). The expression of all the genes was identical in the fore and hind wings. The expression pattern of these genes has been studied in wing discs of other hymenopterans. The expression of these genes in honey bee is also conserved as in other hymenopterans (Abouheif *et. al.,* 2008). These genes however show differential expression between the wing and haltere of *Drosophila*- suggesting that in the dipteran lineage some of the common wing patterning genes have been targeted by Ubx.

Cut protein could be detected along the D-V boundary of the fore and hind wing in *Apis mellifera* in the fifth larval instar. The specificity of staining was demonstrated by the co localisation of DAPI (staining the nucleus) and the protein was clearly localised to the nuclei of the boundary cells.

Spalt was detected in a broad stripe along the A- P boundary of fore and hind wing in *Apis mellifera* in fifth larval instar. Once again the specificity was demonstrated using DAPI staining.

Engrailed was detected in the posterior compartment of the wing disc while Extradenticle was detected in the proximal region of wing disc.

The conservation of expression pattern of these genes with those in other hymenopterans and dipterans suggested that the wing-patterning network has remained identical in all insects including honeybee.

2.13: Generation of polyclonal antibodies against honeybee Ubx

In order to generate specific antibodies against Ubx of *Apis mellifera*, a recombinant protein containing only N- terminal Ubx was used as antigen (as described in materials & methods). **Plate 2.14** shows schematically the region of Ubx used for generation of antibody. The figure also shows the purification of protein using Ni-NTA column. The antibody obtained was used to probe embryonal lysate from both *Apis mellifera* and *Drosophila melanogaster*. The antibody was specific to Ubx of *Apis* and couldn't detect the Ubx of *Drosophila* by Western blot hybridization. This suggested that the antibody was specific to Ubx of honey bee.

Antibody staining was carried out using anti Ubx (*Apis*) on fore and hind wing bud of honey bee and signal was detected using DAB kit. There was a higher concentration of Ubx in the hind wing as compared to the fore wing. Antibody staining was repeated using fluorescent secondary antibodies. Ubx protein was visible in both the fore and hind wing. Unlike *Drosophila melanogaster* and *Precis coenia* where Ubx expression is limited only to the peripodial membrane of the fore wing disc; in *Apis*

mellifera Ubx expression was found in the complete fore wing disc. The peripodial membrane couldn't be detected in the insect (**plate 2.15**).

2.14: Determination of the larval stage of honey bee at which ChIP should be carried out

As already described, Cut protein expression served as a marker to identify the larval stage of honeybee, which is equivalent to the late third instar of *Drosophila melanogaster*. Cut expression was detected using antibodies against *Drosophila* Cut protein and fifth larval instar of honeybee was identified as the stage which is equivalent to the third larval instar of *Drosophila* and ChIP was carried out on tissues isolated at this stage.

2.15: Antibody Purification

In order to carry out ChIP experiments, it was necessary to purify the IgG fraction of the whole sera so that any unwanted interactions between the chromatin and sera may be reduced. The process of using IgG has also given reproducible results as the amount of IgG in antibody and pre immune IgG can be quantified and equal amounts of both can be used.

IgG fraction was purified from both the pre immune and anti Ubx fractions and used further in the study (**plate 2.16**).

2.16:Chromatin Immunoprecipitation

As discussed in the introduction, success of ChIP experiment depends on a variety of factors, few of them have been discussed earlier. Right from collecting tissue at the correct stage and their fixation, choice of antibodies and correct time for sonication have deep impacts on the result that is obtained by ChIP.

Fixative to be used

In order to carry out chromatin immunoprecipitation, fixation had to be carried out using 1.5 % formaldehyde. In general, while performing ChIP on transcription factors, a higher concentration of formaldehyde is used (unlike 1% for histone proteins or DNA damage repair proteins- which are associated more tightly with the DNA). Also, freshly prepared formaldehyde (from Paraformadehyde) was used. This was necessary as commercial formaldehyde contains methanol for stabilising the formaldehyde. The presence of methanol can mask the epitiopes of the antigen (Ubx protein in this case), making it difficult for the antibody to recognise and bind the protein in solution. Also, it is very necessary that the formaldehyde that is prepared be stored at room temperature, in dark for a period no longer than 15 days (or until precipitate can be seen- whichever is earlier). Storing formaldehyde in cold leads to its precipitation and so does its storage in light. Use of such formaldehyde may lead to incomplete fixation which might disrupt protein DNA interactions during sonication and hence a good amount of DNA may never be recovered.

Antibody to be used

One of the most important factors that affect the success of the ChIP experiment is the antibody that is used for the experiment. The antibody should not only be polyclonal which can recognise the protein under various states but should also be very specific. The design of antibody in the current study reduced the chances of any non-specificity and hence enhanced the success of ChIP experiment.

However, all antibodies have some amount of affinity for any general antigens. In order to rule out the possibility of any general (non specific interactions) binding leading to false peaks, ChIP was always carried out on the same chromatin on pre immune IgG and anti Ubx IgG. Any enrichment that would be found in both the experimental datasets can be considered noise while those specific to anti Ubx IgG should be chosen in all cases. To further minimise the chances of false positives, same amounts of pre immune IgG and anti Ubx IgG were used.

Standardisation of sonication time

It is essential that the chromatin must be sheared to size of 300- 500 bp post fixation, so that one may study the DNA protein interactions in vivo with ease. It is important however that this shearing is done under very optimum conditions and minimum shearing force is applied so that the DNA- protein interactions are not disrupted. Hence, sonication of chromatin was always carried out in ice cold water and various conditions were tested to check the minimum time required for attaining the desired size of DNA. In order to standardise the time of sonication (essentially the minimum time required for sonicating the chromatin to around 300- 500 bp), the discs were sonicated in identical conditions for 5, 10, 20 and 35 minutes. When the sonication was carried out for 5 minutes, most of the DNA was seen near the well (indicative of the fact that shearing was quite in sufficient and mostly DNA was of high molecular weight). When sonication was carried out for 10 minutes, high intensity band was seen around 300 bp- this time was considered as the ideal time for sonicating DNA.

Chromatin was also sonicated for 20 and 35 minutes and it was observed that with increasing sonication time, smaller DNA fragments increased and hence for all the subsequent ChIP experiments, a total sonication time of 10 minutes was used (**Plate 2.17**).

ChIP-Western

Following ChIP, it is essential to determine if the antibody was able to detect the antigen in solution (essentially if the anti Ubx antibody is of ChIP grade or not). A Western blot hybridization of chipped samples was carried out to test this. Two chipped samples- one immunoprecipitated using anti Ubx IgG and one using pre immune IgG were probed by Western blot hybridization using anti Ubx antibody. A band of ~ 35 kDa was visibly enriched in the anti Ubx IgG immunoprecipitated sample as compared to the pre immune IgG immunoprecipitated sample, which suggested that the anti Ubx IgG being used in the study is of ChIP grade and did pull down the Ubx protein preferentially.

Reverse cross linking & elution of DNA

Reverse cross linking of DNA protein complexes is another crucial step. It has been observed that for reverse crosslinking the DNA-protein complexes, chromatin should be incubated for no longer than 7 hours at 65ºC. Storing the DNA for longer periods at this temperature may harm the downstream processing events of sequencing. This may however not apply for DNA that is being analysed by tiling array.

Elution of DNA followed by its precipitation too needs attention when the DNA has to be used for sequencing. Sonicated salmon sperm DNA is usually added as a carrier to enhance the precipitation of DNA but it has again been observed that this DNA can be sequenced during the sequencing process and obscure the reading of the chipped DNA fragments. Hence in the current study, salmon sperm DNA was not used to enhance precipitation.

Another crucial fact that was taken into account was the salt that should be used for precipitation. Sodium chloride was used to precipitate DNA as those ions donot affect the sequencing process in general. Isopropanol precipitation was absolutely avoided as it precipitates a good amount of salt which hinders the process of ligation during library preparation. Also, molecular biology grade glycogen (free of contaminating DNA) was used to precipitate DNA.

Finally, it is essential to use input DNA as the baseline for comparison as it is free of any bias (unlike pre immune IgG chipped DNA) and also takes into account various other artefacts like PCR bias which may arise in particular regions of genome due to selective amplification. Hence for each sample, one input, one pre immune IgG chipped DNA and one anti Ubx IgG chipped DNA was sequenced and compared.

ChIP- Seq was carried out on fore and hind wing and for each experiment input DNA, anti Ubx chipped DNA and pre immune IgG chipped DNA were sequenced. Two replicates of fore wing and two of hind wing were sequenced in all cases.

2.17: Sequencing of chipped DNA & reads obtained

Sequencing of chipped DNA was carried out by GenotypicTechnology [P] Ltd, Bangalore. Library was prepared according to standard protocols of the company. The Bio analyser profiles were then checked and only the libraries that were found suitable were selected for sequencing. Bio analyser profiles of input, anti Ubx and pre immune chipped DNA of various replicates have been shown in **plates 2.18- 2.29**.

Important criteria that needed to be met when selecting the libraries for further sequencing was the length of DNA fragments as detected on Bio-analyser after library preparation. A single peak in the profile indicates that most of the DNA has been sheared around a particular length and has been subsequently pulled down. Also, it is important to see if the DNA peaks around almost the same size as seen on agarose gel after sonication.

As, all the profiles met the necessary criteria they were used for sequencing using Illumina Sequencer.

Short 36 base pair reads were obtained from the company for various samples of chipped DNA which were subsequently used for data analysis. On an average around 30 million reads were obtained for each sample from a particular run. These reads were subsequently used for data analysis which has been described in detail in the next chapter.

Summary

To summarise, the chapter has described the following aspects:

1. Expression of wing patterning genes is conserved in honeybees and other insects and major wing patterning genes donot show any difference in expression between fore and hind wing.
2. Early fifth larval instar of honey bee has been identified as the correct stage for carrying out chromatin immunoprecipitation experiments – the stage that corresponds to late third instar of fruit fly.
3. Polyclonal antibodies specific to Ubx of honeybee were raised which were found to be ChIP grade in ChIP-Western studies.
4. Chromatin immunoprecipitation was carried out which was followed by sequencing and a good number of reads for each of the samples were obtained.

In the following chapters, analysis of the reads has been described in terms of identification of peaks associated with these reads (the regions bound by Ubx). These regions bound by Ubx have been further analysed and compared with Ubx bound regions in *Drosophila* in terms of the genes they are associated with and the motifs they contain.

Plates

Chapter 2

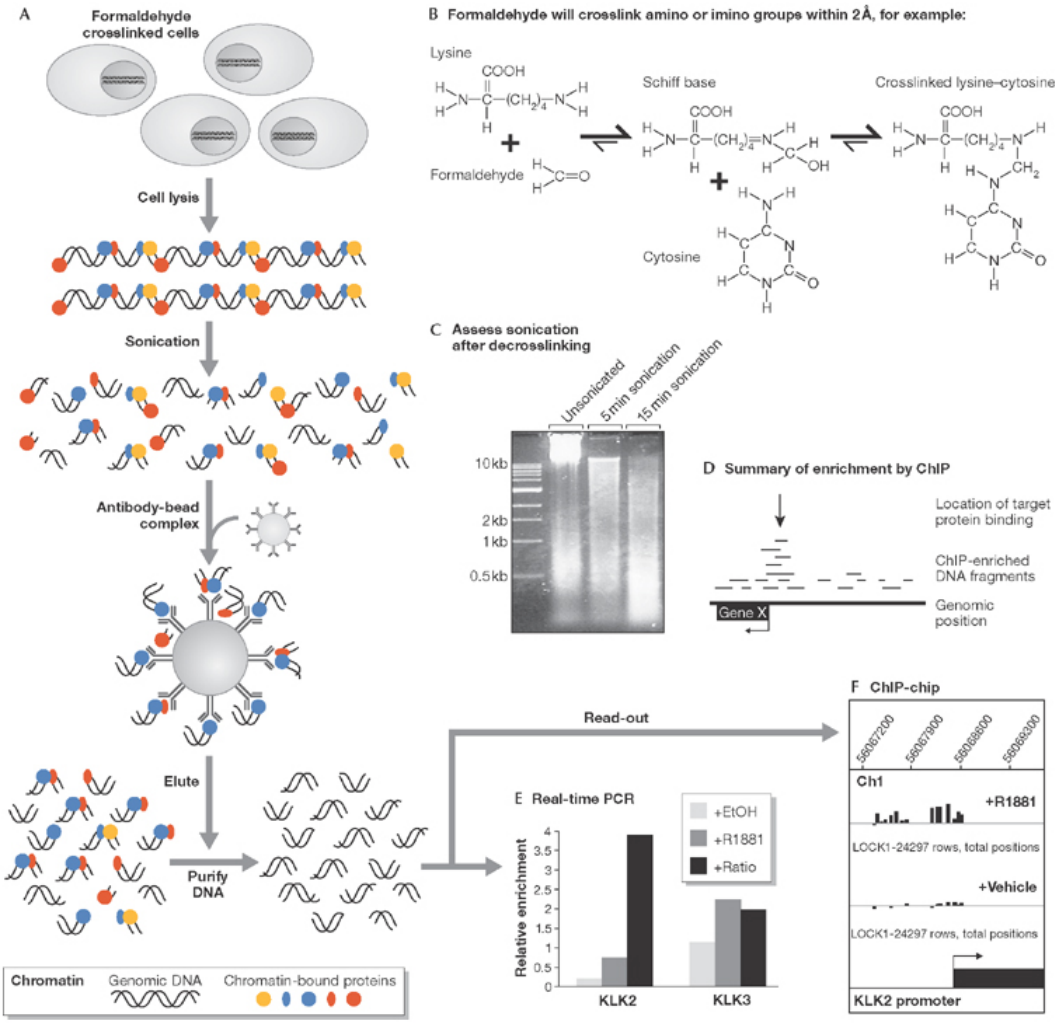


Plate 2.1: An overview of ChIP protocol (Massie *et. al.*, 2008)

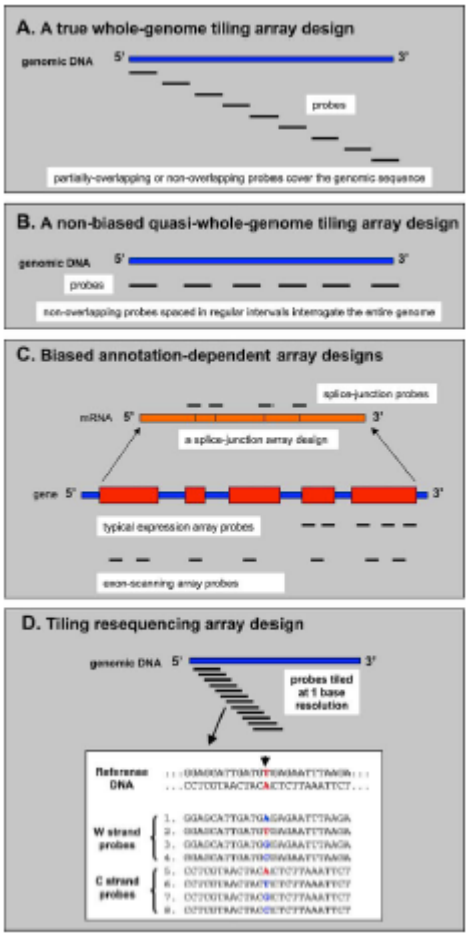


Plate 2.2: Various types of tiling arrays.

A & B: Unbiased whole genome tiling arrays representing the whole genome having overlapping and non overlapping probes respectively.

C: Biased whole genome tiling arrays representing exon junctions, splice sites etc.

D: Tiling resequence array: each nucleotide of the reference genome is represented with eight set of nucleotide probes (Mockler *et. al.*, 2005)

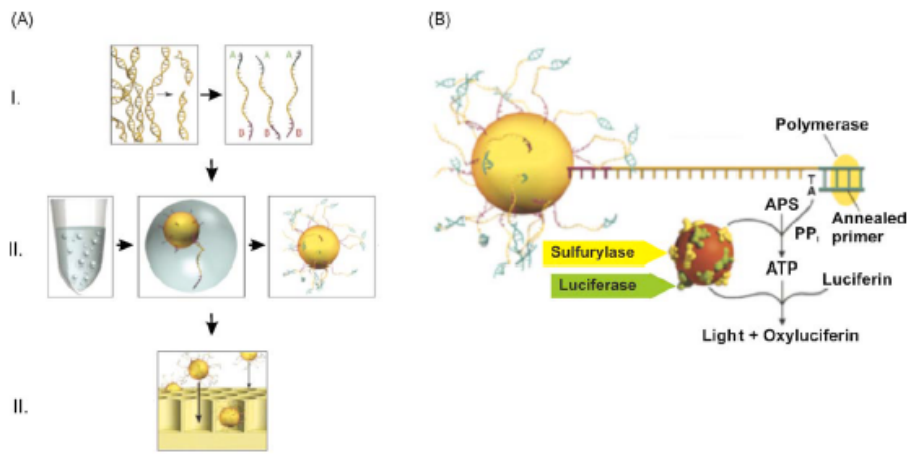


Plate 2.3: Workflow for 435 Genome Sequencer

A: Library Preparation

i) Specific primers are ligated to DNA sequences

ii) DNA is amplified on the beads

iii) The beads are then loaded on microtire plates

B: Sequencing

Pyrophosphate reaction that is used by the sequencer to report the incorporation of nucleotide .

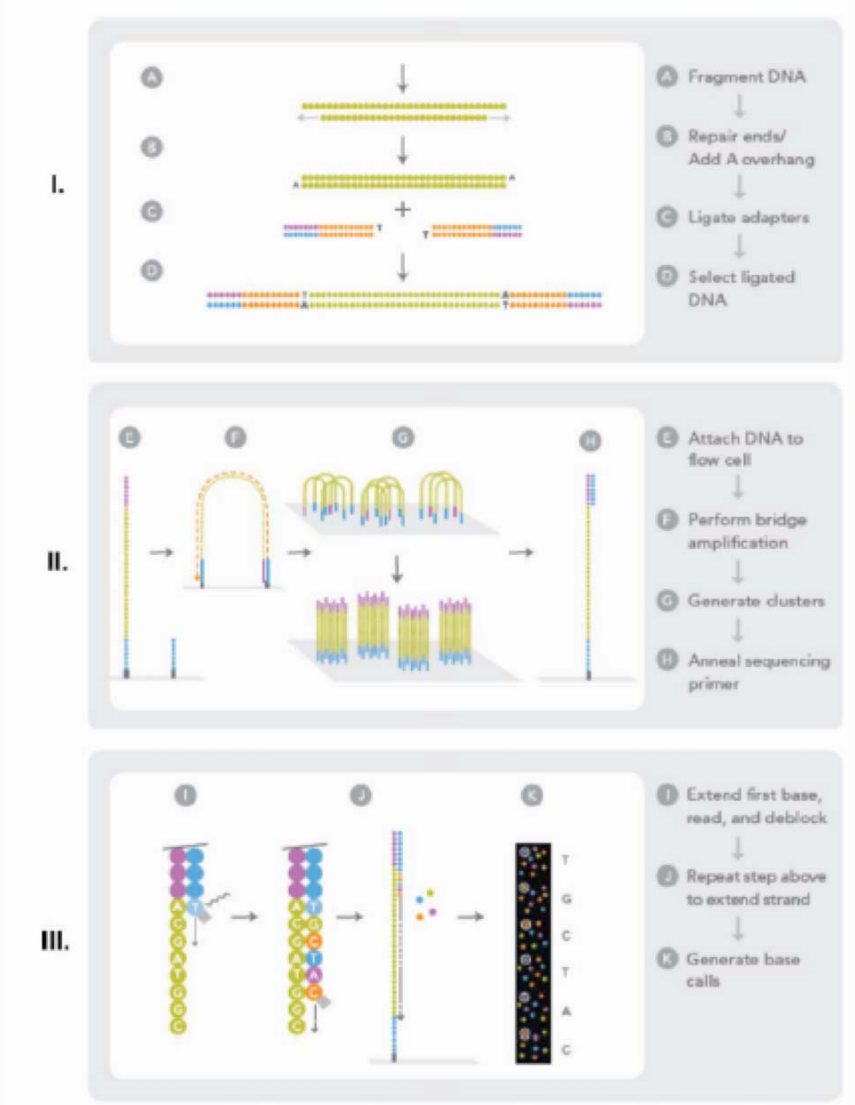


Plate 2.4: Work flow of sequencing by Illumina sequencer

I. Library is prepared by ligating the illumine specific primers to DNA ends

II. The attached DNA fragments are then applied to the glass capillary where the DNA forms a bridge which are subsequently amplified

III. The amplified fragments are subjected to sequencing using nucleotides

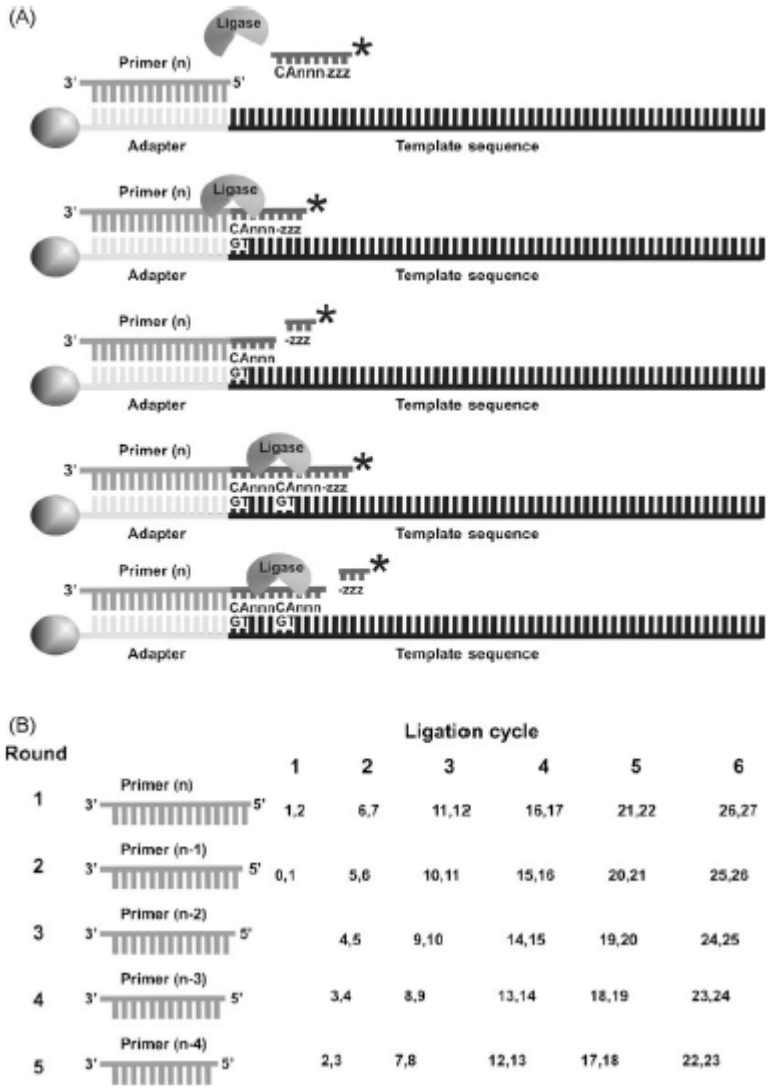


Plate 2.5: Over view of sequencing by ligation reaction used by ABI SOLiD system.

Plate 2.6: Different locations from where the honey bee was collected.

A. A view of farmland in Kolhar with a rich flora (sunflower) that provides a good amount of pollen and nectar for honeybees
B. A view of the boxes which are usually kept close to the field adjacent to a water body by farmers
C. Bee boxes/ hives maintained at Heinrich Heine University, Dusseldorf, Germany where food was artificially provided
D. The flora in this part of the world remained a bit low but the artificial diet was sufficient for honeybees.

Plate 2.7: Maintenance of honeybees at IISER, Pune

A. Two hives within a fenced area

B. A single hive seen with a top lid covered with aluminium sheet (to protect the hive from rains)

C. The same hive seen after the removal of top cover. Jute cover helps in maintaining a constant temperature within the hive.

D. The hive with bees seen after removal of jute cover. The plastic bag seen contains sugar syrup. Small holes are made in it from which sugar drips slowly and is taken by bees

E. A close view of the frames of hive showing bees

F. A still closer view of frame (pl notice the spacer wood at the side- which is used to maintain a constant distance between the frames for the bees to move between them)

Plate 2.8: Life of honeybees inside the hive

A. A single frame of the hive showing nurse, worker and queen bees

B. Queen bee shown with an arrow and worker bee entering the hive.

C. A closer view of the frame

D. Another view of the frame showing various stages of the honey bee

E. A closer view of the hive showing the capped pupae and the feeding larvae

F. The eggs have been labelled with white arrows in this view.

G. Another frame showing some collection of pollen (yellow in colour marked with white arrows)

H. Honey being stored in the chambers.

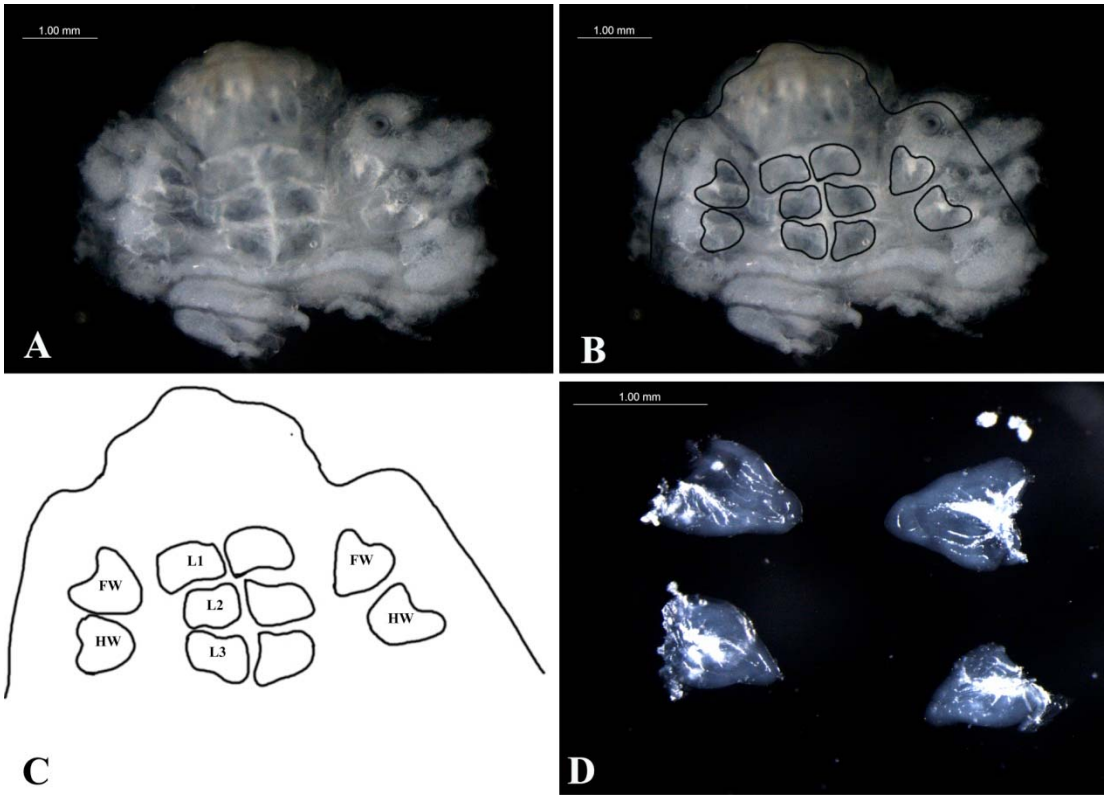


Plate 2.9: Location of wing and leg buds in the honey bee larva

A. Anterior portion of honeybee larva after partial dissection as viewed from ventral side

B. A rough outline of the wing and leg discs made on the larval body

C. The rough outline of body and buds

D. The wing buds arranged and viewed under a stereo microscope (the wing buds at top are the fore wing buds while those at bottom are hind wing buds)

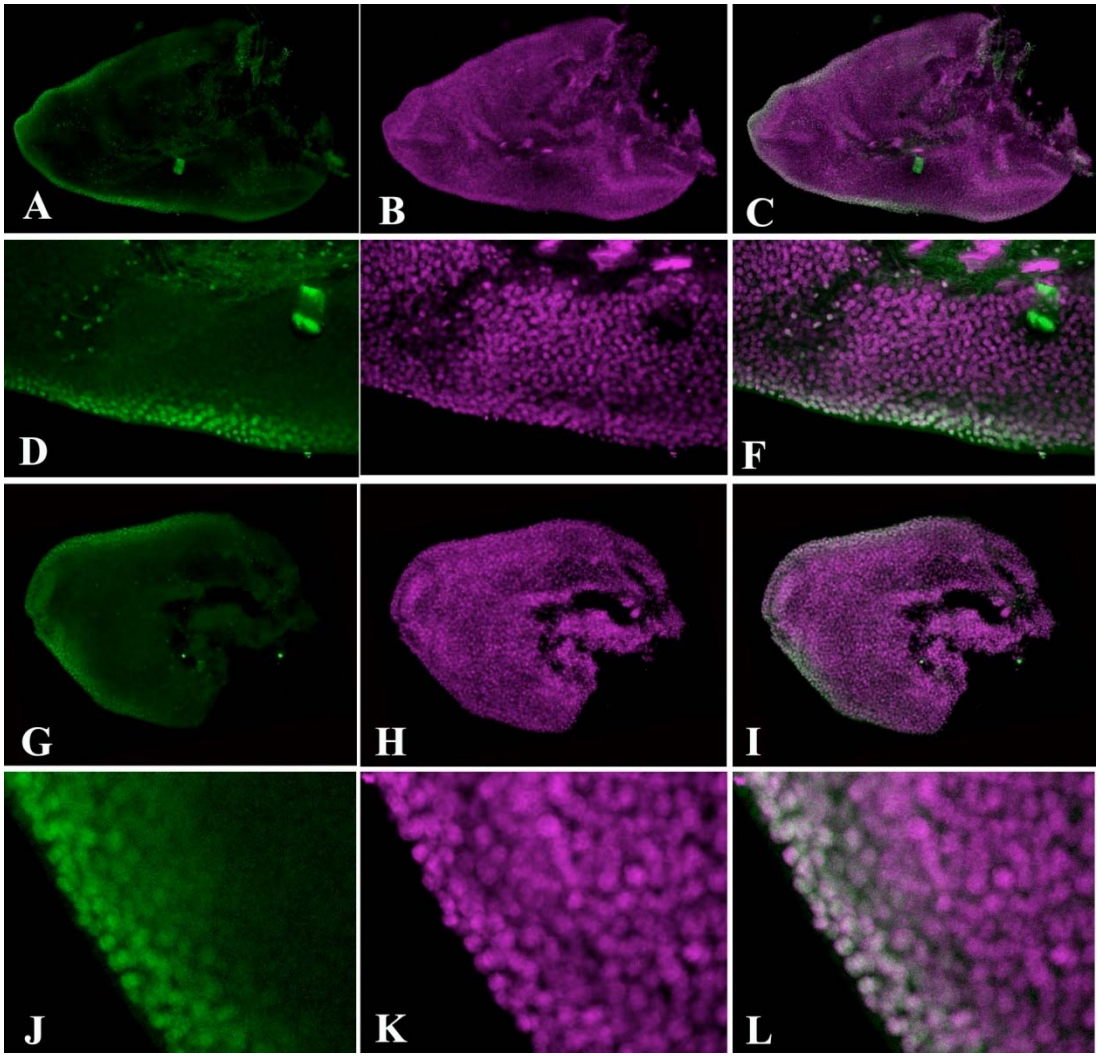


Plate 2.10: Antibody staining of wing buds of Apis mellifera using anti cut antibody

Panels A-F are fore wing while G-I are the hind wing images

A, D, G and J show anti cut staining

B, E, H and K show the nucleus stained with DAPI

C, F, I and L show the merged image of DAPI and anti cut.

Images of panel A, B, C , G, H and I have been taken at lower magnification (25X) while those of D, E, F, J, K and L have been taken at a higher magnification (40X)

(Distal end is towards left and proximal end is towards right in all the images)

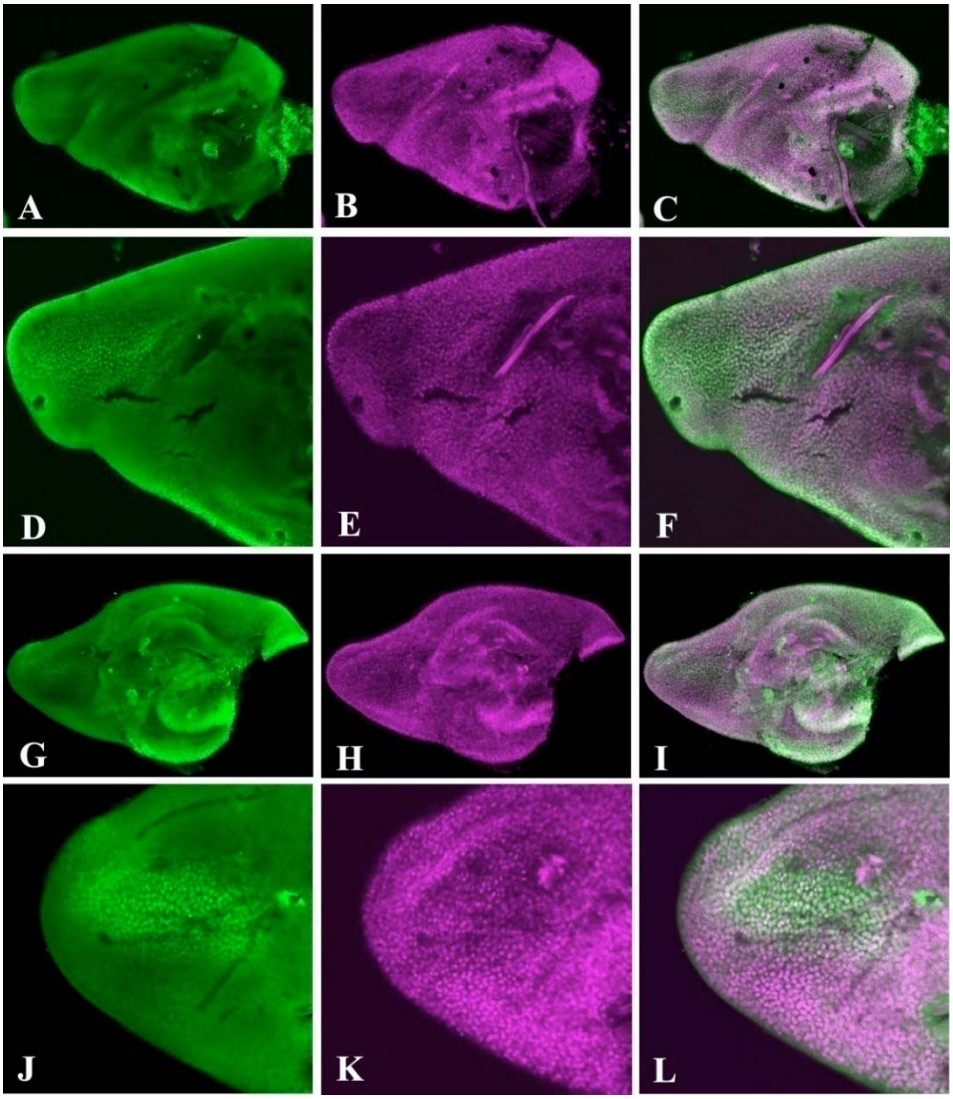


Plate 2.11: Antibody staining of wing buds of Apis mellifera using anti spalt antibody

Panels A-F are fore wing while G-I are the hind wing images

A, D, G and J show anti spalt staining

B, E, H and K show the nucleus stained with DAPI

C, F, I and L show the merged image of DAPI and anti spalt.

Images of panel A, B, C , G, H and I have been taken at lower magnification (10X) while those of D, E, F, J, K and L have been taken at a higher magnification (25X)

(Distal end is towards left and proximal end is towards right in all the images)

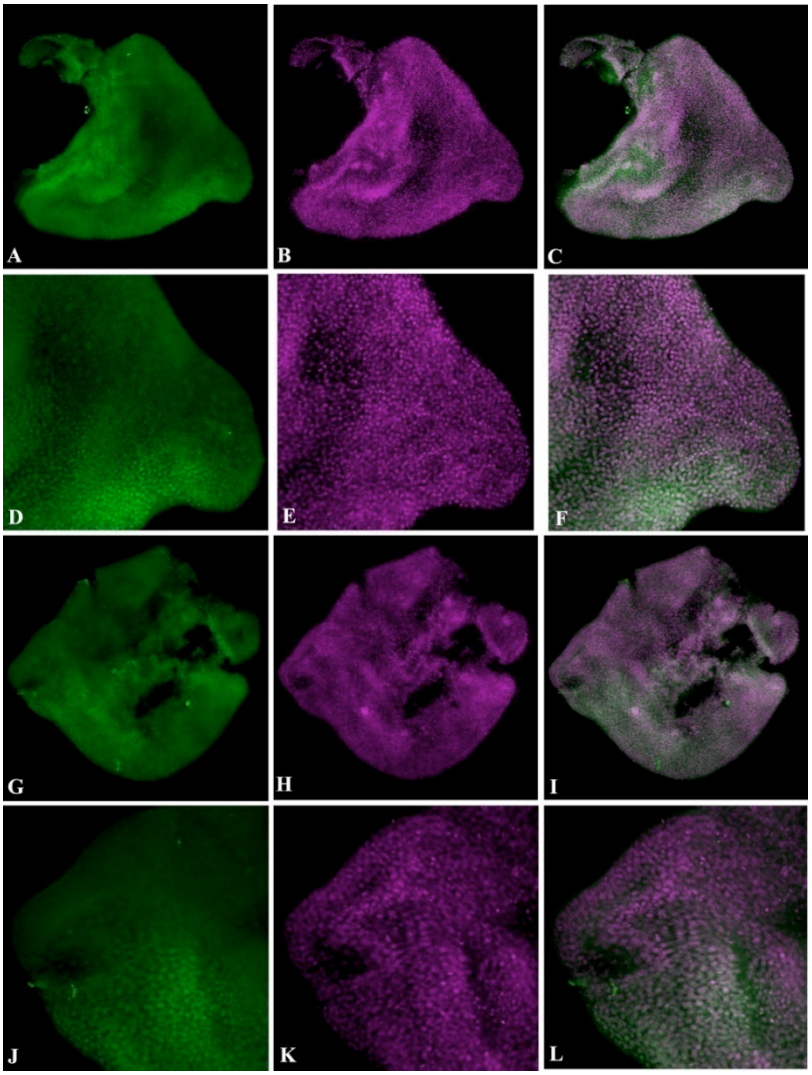


Plate 2.12: Antibody staining of wing buds of Apis mellifera using anti engrailed antibody

Panels A-F are fore wing while G-I are the hind wing images

A, D, G and J show anti engrailed staining

B, E, H and K show the nucleus stained with DAPI

C, F, I and L show the merged image of DAPI and anti engrailed.

Images of panel A, B, C , G, H and I have been taken at lower magnification (10X) while those of D, E, F, J, K and L have been taken at a higher magnification (25X)

(Anterior is towards top and posterior is towards bottom in all the images)

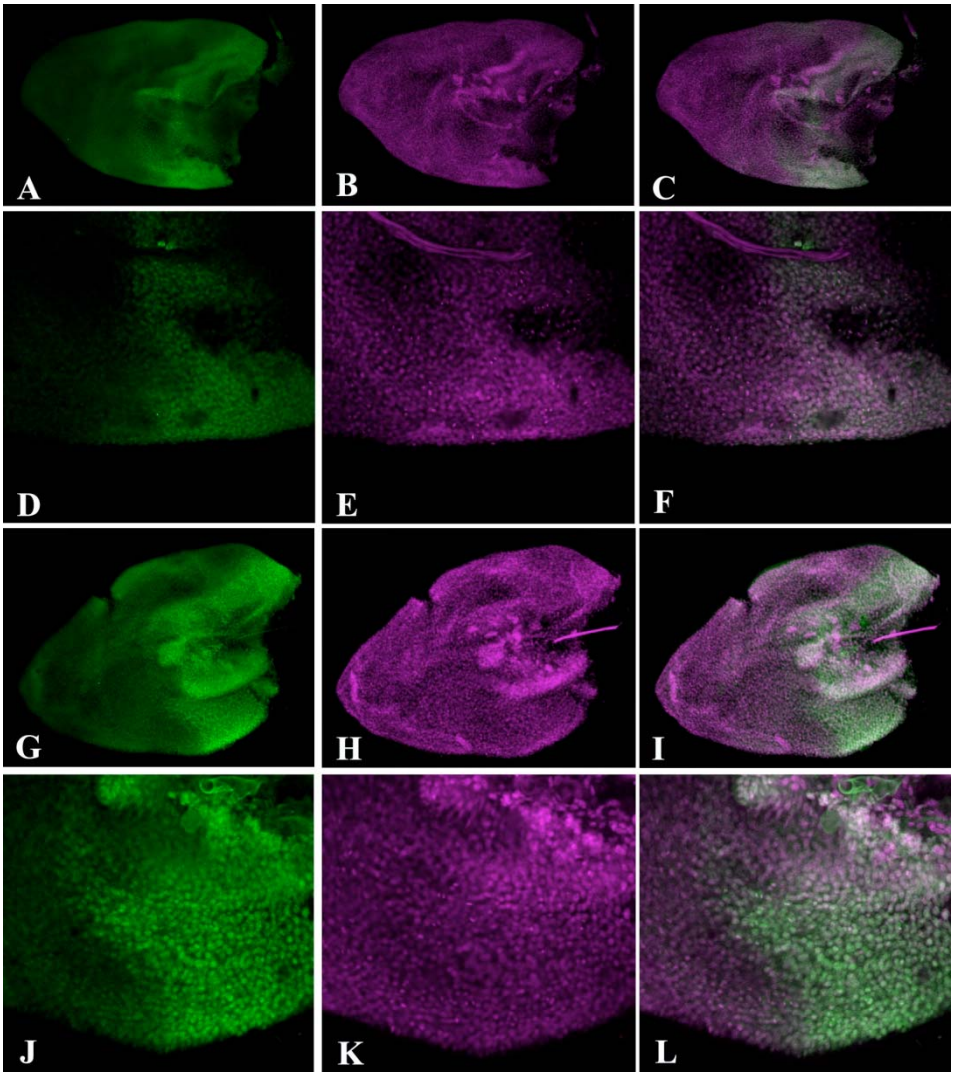


Plate 2.13: Antibody staining of wing buds of *Apis mellifera* using anti extradenticle antibody

Panels A-F are fore wing while G-I are the hind wing images

A, D, G and J show anti extradenticle staining

B, E, H and K show the nucleus stained with DAPI

C, F, I and L show the merged image of DAPI and anti extradenticle.

Images of panel A, B, C , G, H and I have been taken at lower magnification (10X) while those of D, E, F, J, K and L have been taken at a higher magnification (25X)

(Distal end is towards left and proximal end is towards right in all the images)

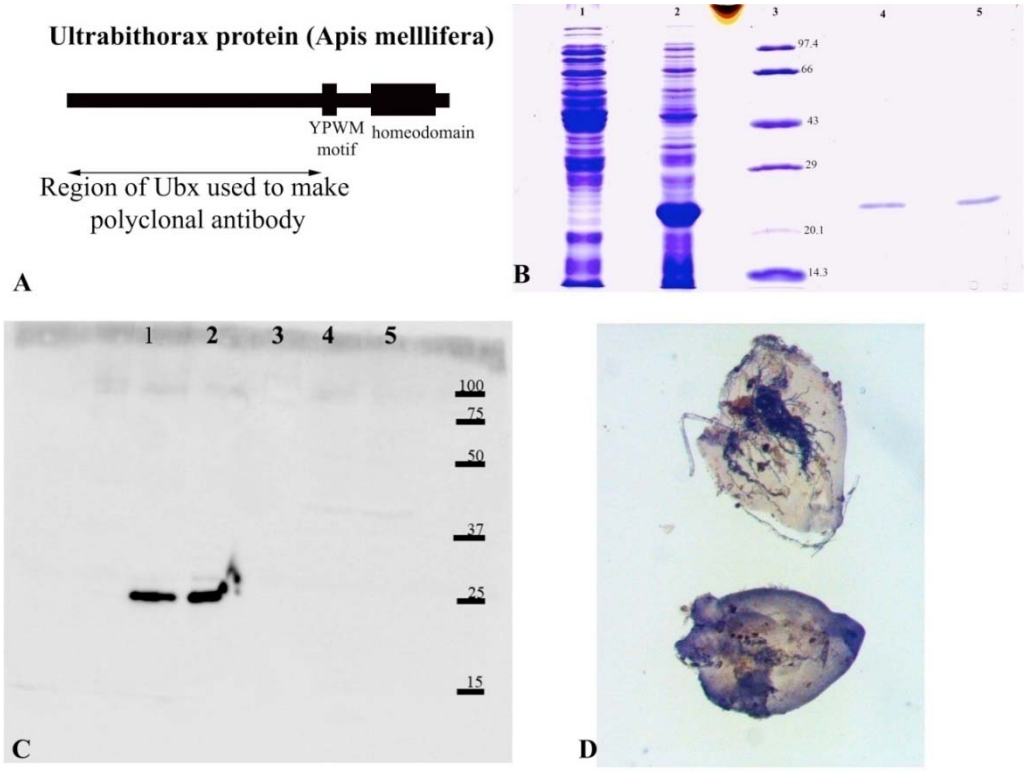


Plate 2.14: Generation of anti Ubx antibody and testing its specificity

A. Schematic representation of Ubx protein used to make antibody

B. Purification of his tagged N- terminal Ubx protein from bacteria

Lane 1: Uninduced bacterial lysate

Lane 2: Bacterial lysate after induction

Lane 3: Molecular weight marker

Lane 4 & 5: Purified protein

C. A western blot probed with anti Ubx (Apis) to demonstrate the specificity of the antibody

Lane 1 & 2: *Apis mellifera* embryonal lysate

Lane 3: Molecular weight marker

Lane 4 & 5: *Drosophila melanogaster* embryonal lysate

D. Antibody staining on the fore and hind wing bud of *Apis mellifera* using anti Ubx antibody demonstrating the differential expression of Ubx between the wing buds.

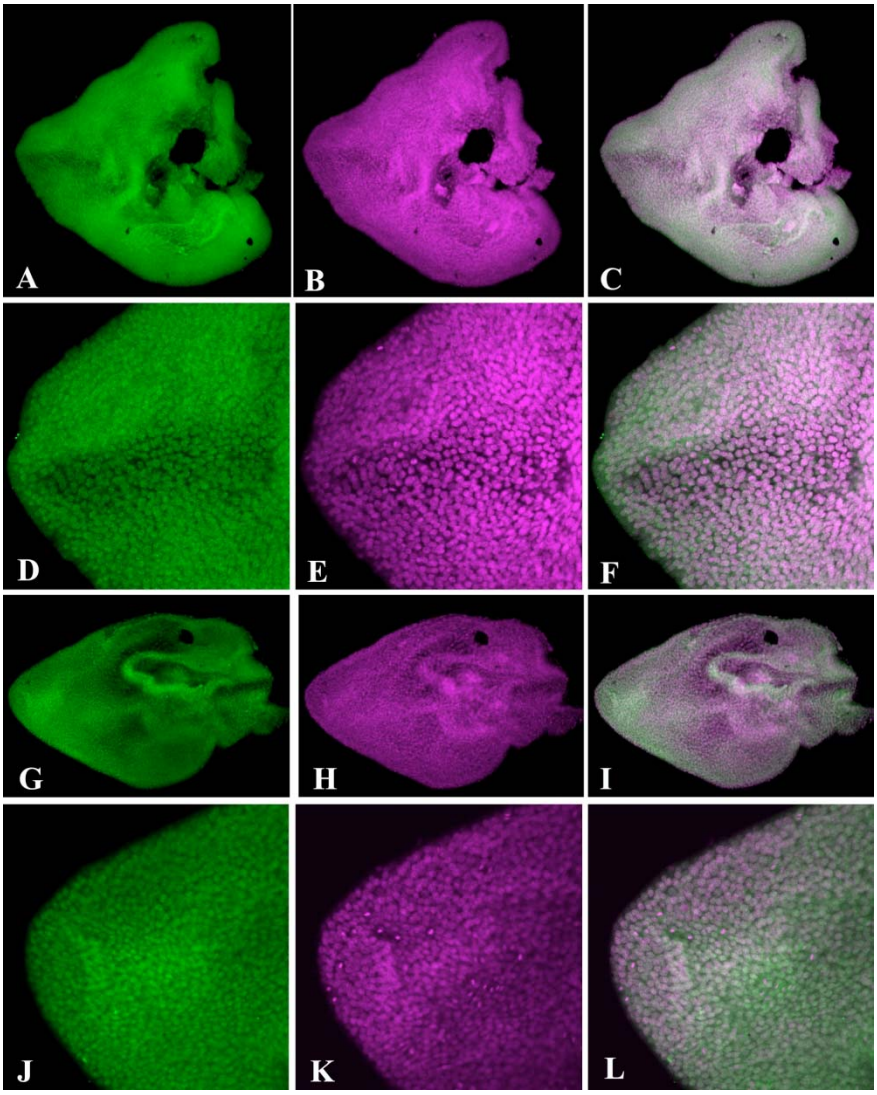


Plate 2.15: Antibody staining of the wing buds of Apis mellifera using anti Ubx antibody

Panels A-F are forewing while G-I are the hind wing images

A, D, G and J show anti Ubx staining

B, E, H and K show the nucleus stained with DAPI

C, F, I and L show the merged image of DAPI and anti Ubx.

Images of panel A, B, C , G, H and I have been taken at lower magnification while those of D, E, F, J, K and L have been taken at a higher magnification

(Distal end is towards left and proximal end is towards right in all the images)

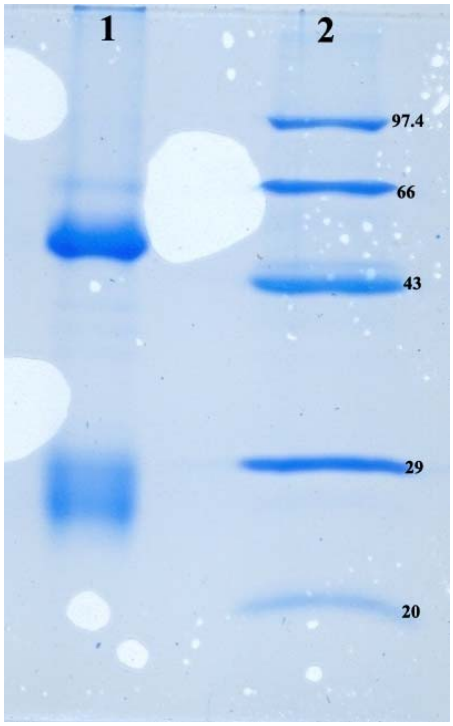
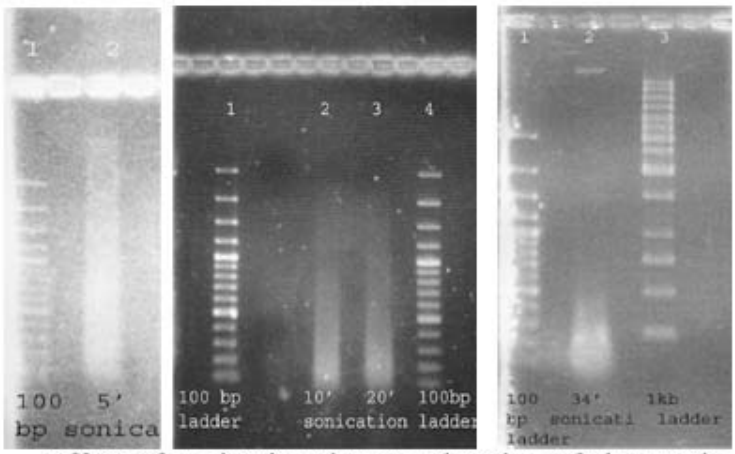


Plate 2.16: IgG purified anti Ubx antibody (as seen on SDS gel)

Lane 1 shows the purified antibody while lane 2 shows molecular weight marker (the molecular weights indicated are in kDa)



Effect of sonication time on shearing of chromatin

Plate 2.17: Standardisation of time required for chromatin sonication

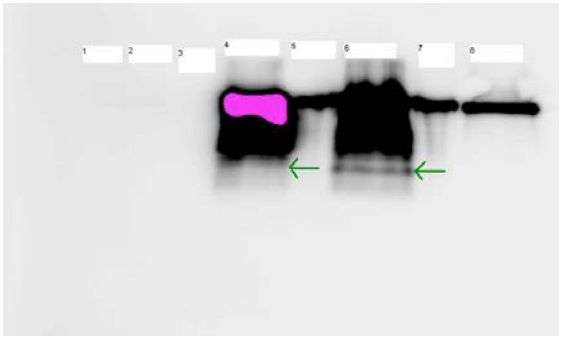


Plate 2.17a: Western blot showing the pull down of Ubx protein in the anti Ubx IgG pull down fraction as compared to pre immune IgG pull down fraction

Lane 1: marker

Lane 2 : input (this is the chromatin which has been pre cleared by incubation with protein a sepharose)

Lane 4: pre immune sera pull down (this lane has been loaded with the pull down products using pre immune IgG)

Lane 6: anti ubx pull down (this lane has been loaded with the pull down products using anti Ubx IgG)

Lane 8: spent lysate (this is the lysate that was left after the binding of anti ubx-protein a sepharose complex to chromatin)

Lane 3, 5 and 7 were not loaded with anything but some flow from adjoining lanes can be seen on western blot

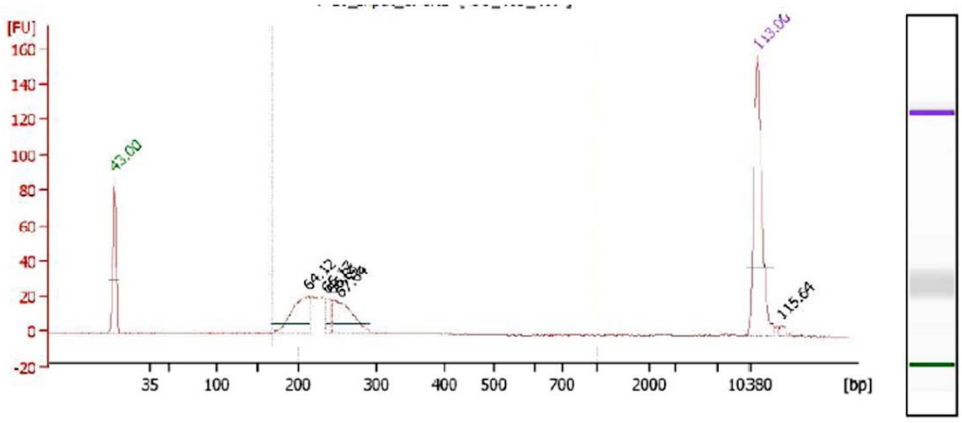


Plate 2.18: Bio analyser profile of Fore wing 10 input DNA

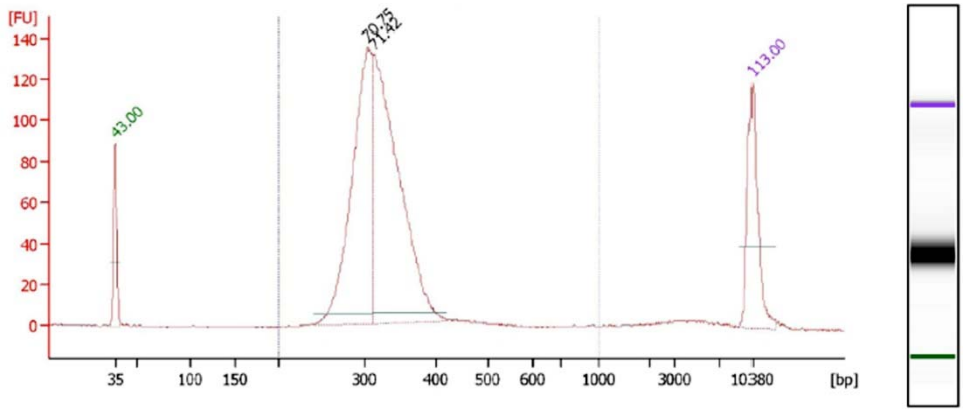


Plate 2.19: Bio analyser profile of Fore wing 8 input DNA

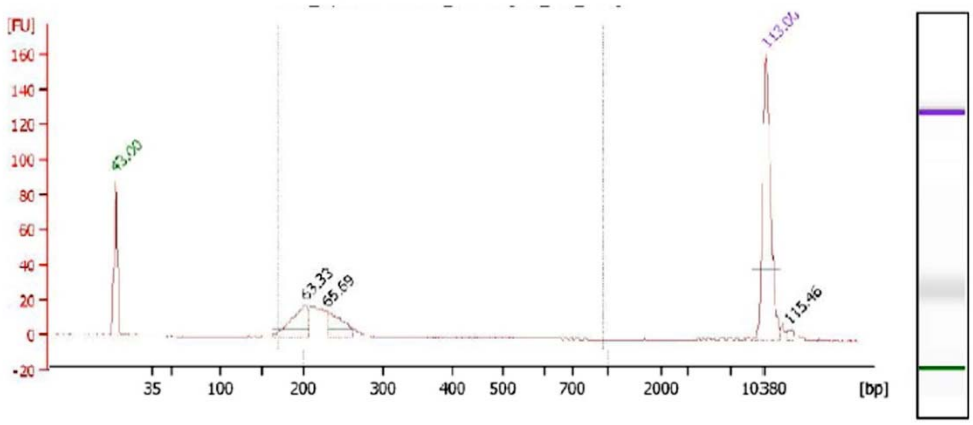


Plate 2.20: Bio analyser profile of Fore wing 10 anti Ubx chipped DNA

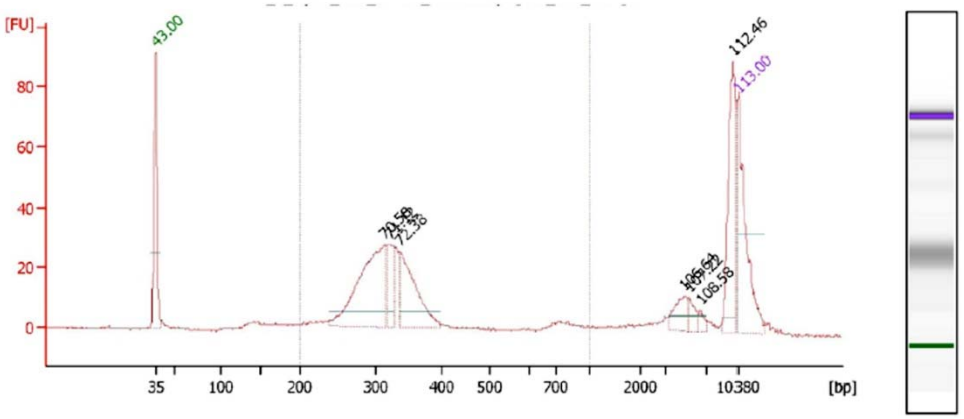


Plate 2.21: Bio analyser profile of Fore wing 8 anti Ubx chipped DNA

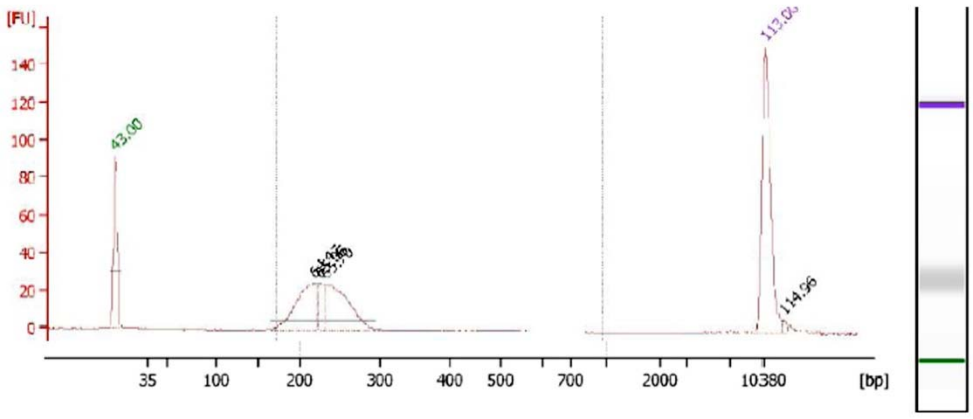


Plate 2.22: Bio analyser profile of Fore wing 10 pre immune IgG pull down DNA

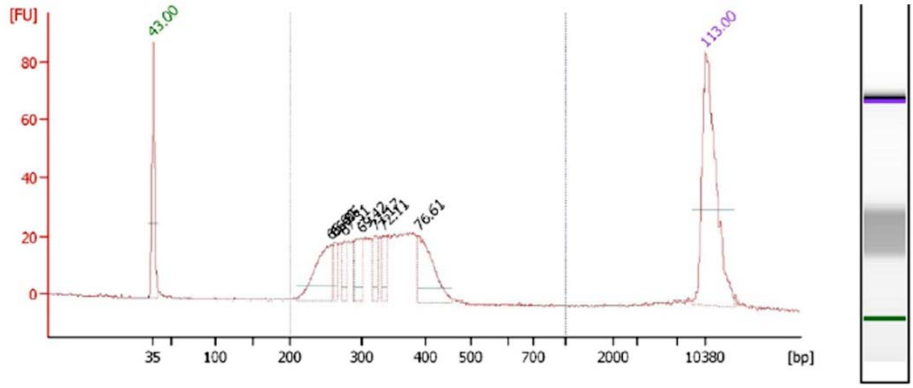


Plate 2.23: Bio analyser profile of fore wing 8 pre immune IgG pull down DNA

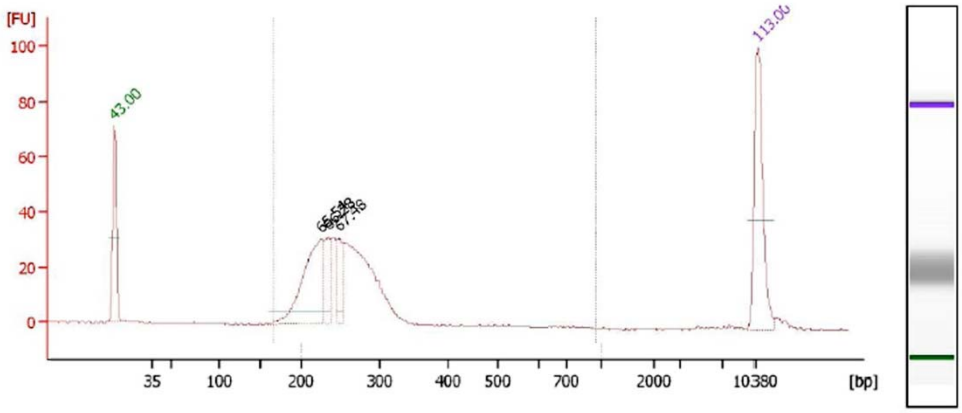


Plate 2.24: Bio analyser profile of hind wing 10 input DNA

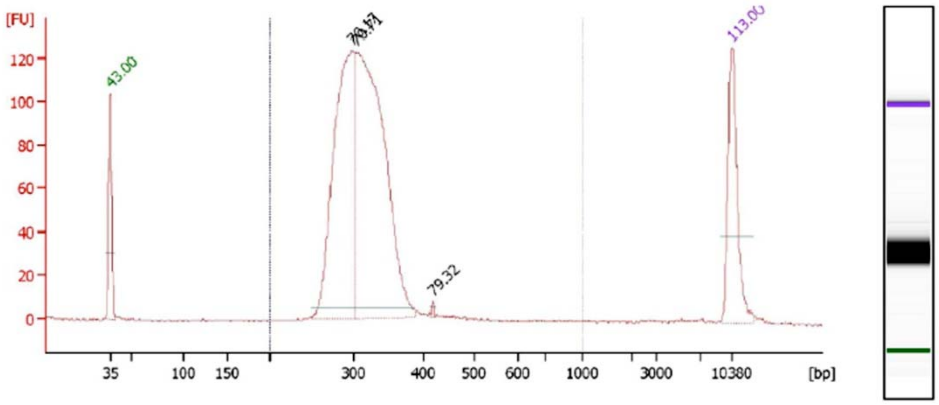


Plate 2.25: Bio analyser profile of hind wing 8 input DNA

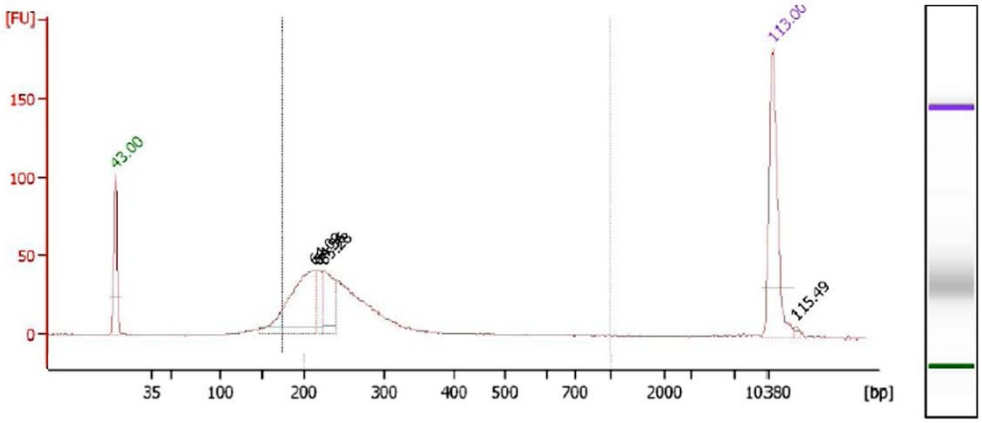


Plate 2.26: Bio analyser profile of hind wing 10 anti Ubx chipped DNA

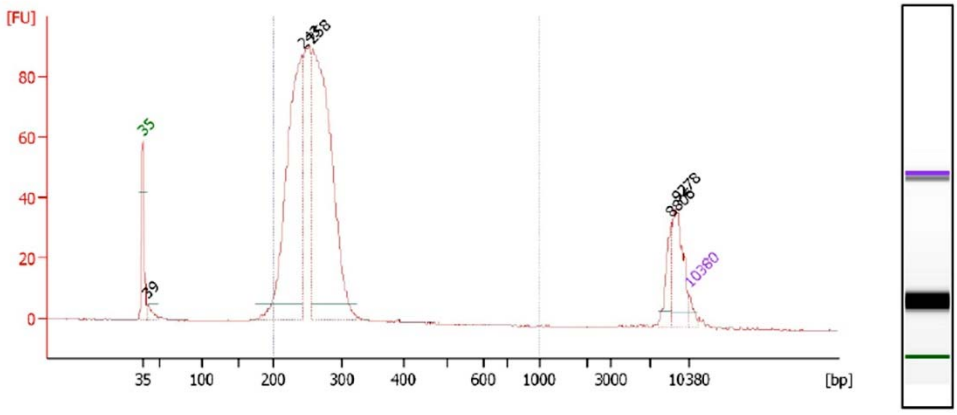


Plate 2.27: Bio analyser profile of hind wing 8 anti Ubx chipped DNA

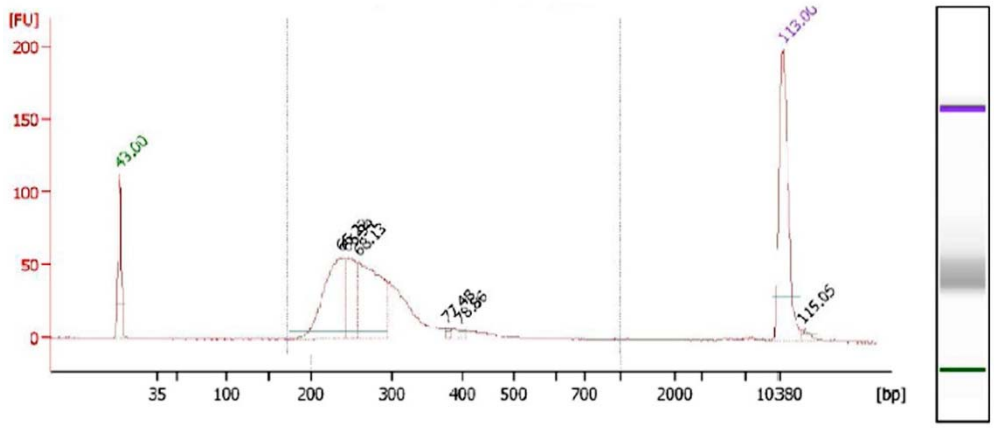


Plate 2.28: Bio analyser profile of hind wing 10 pre immune IgG pull down DNA

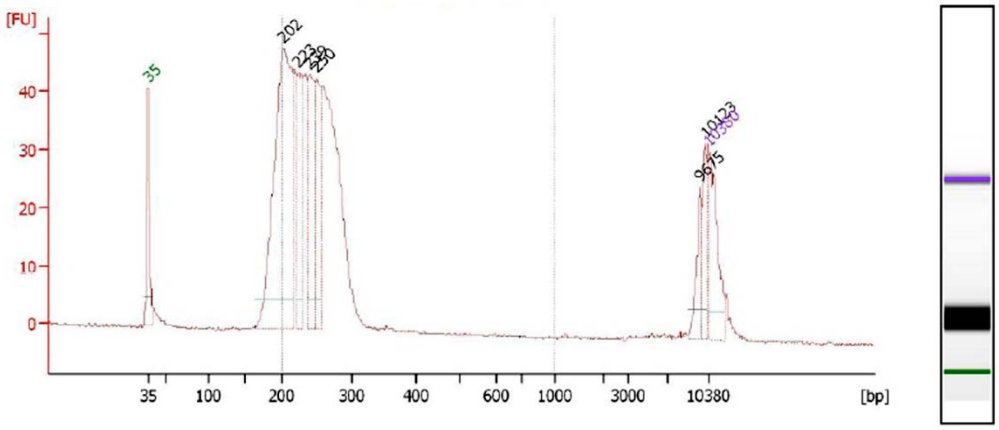


Plate 2.29: Bio analyser profile of hind wing 8 pre immune IgG pull down DNA

Chapter 3

Analysis of the chipped DNA sequences

Introduction

In the post genomic era, generation of enormous amounts of data has necessitated the development and use of various computational algorithms that can efficiently analyse and manage these datasets. Various tools in bioinformatics have therefore become very useful and indispensable while dealing with such data. Chromatin Immunoprecipitation (ChIP) results in the enrichment of DNA fragments bound to the transcription factor of interest. The sequencing of these DNA fragments would therefore yield sequences that are identical to the whole genome of the organism being studied but these sequences should map to fewer locations (peaks/ regions bound by the transcription factor/ the enriched DNA fragments). Prior to such alignment, it is very important to check the DNA sequences for their quality. Only the high quality sequences with a good number of reads should be subjected to further analysis. Once the chipped DNA is subjected to sequencing, the sequence of DNA is obtained as large files with small reads. Typical sequencers give around 36 bp long reads as the output usually the number of such reads are in the range 30-50 million. A number of parameters have to be looked at while testing the quality of the sequences, which will be discussed in detail later. It is essential to see if the data that is obtained post sequencing is free of any PCR bias or any type of cross contamination.

The sequences need to be aligned to the genome so that their enrichment at any particular location may be detected subsequently. More sequences align to a particular locus and form a summit, which is detected as a peak or enrichment later.

This chapter describes various parameters for testing the quality of DNA sequences. Various peak calling algorithms, which can be used after alignment of the DNA sequences to the genome have also been discussed.

3.1: Peak Calling algorithms

The enrichments or the peaks are detected by peak calling softwares. The peak calling algorithms are essentially of multiple types- the window based approach, which first defines the start and end of a candidate region and then count the number of reads within it; overlap based approach which first identifies the peaks on the basis of local maxima of read-overlaps and then set the corresponding start and end of peak region; and the third is based on hidden markov model which describes the read accumulation along the genome as a sequence of two different states- binding sites and background.

The ChIP experiments themselves are of two types- one including a negative control while the second is without it. When a negative control is available, statistical significance of the peak can be calculated on the basis of background. Some algorithms are designed to work only when a control is available while some do not incorporate control data. Some algorithms, however accept both forms- of chip data- with and without control.

In order to evaluate the performance of nine publically available peak calling algorithms – namely PeakFinder, GeneTrack, FindPeaks, SISSRs, QuEST, MACS, CisGenome, PeakSeq and Hpeak; a study was carried out by Laajala *et.al,*. Four different datasets were used in the study, which measures the binding of three human transcription factors namely STAT6, NSRF and FOXA1 from various studies. All the peak calling algorithms were used with their default parameters in this study. It was found that all the peak calling algorithms identified the same region as the peak but the length of the peak that was reported was quite variable. PeakFinder, PeakSeq and Hpeak identified longer peaks while MACS, FindPeaks, SISSRs and CisGenome reported shorter peaks. GeneTrack and QuEST identified only single co-ordinates without boundary estimates (Result summarised in **plate 3.1**).

The number of peaks identified by various peak-identifying algorithms varied greatly. PeakFinder and QuEST identified the smallest number of peaks while FindPeaks,

SISSRs and MACS identified a much larger number of peaks. GeneTrack, CisGenome and Hpeak showed enormous variability in the number of the peaks they detected between samples. On investigating the average overlap of peaks obtained from one algorithm with those of others it was found that the peaks found by algorithms reporting the shortest number of peaks was also found by other algorithms (Result summarised in **plate 3.2**).

When looked for reproducibility of detections across two different ChIPseq experiments, best reproducibility was found with PeakFinderC and QuEST (QuEST gave only single coordinate peaks while PeakFinder gave extremely long peaks). The results with any algorithms improved when negative control was used with the experiments. PeakFinder and QuEST identified a small number of peaks, which remained consistent across samples. While considering the top 1000 or top 2000 peaks, all the peak finding algorithms except SISSRsC showed very high reproducibility (Result summarised in **plate 3.3**)

qPCR was subsequently used in the study to evaluate the peaks called by various peak calling algorithms. Three different datasets- one from NSRF, second from FOXA1 and third from STAT6 were used. For NSRF dataset, all the peak calling algorithms proved to be the identical in identifying the peaks validated by PCR, except QuEST which showed somewhat better results. For FOXA1 dataset, MACS had a clear advantage. For the STAT6 dataset, FindPeaks and QuEST seemed to be best and MACS performance was average(**Plate 3.4-3.6**)

Out of the few algorithms that were compared in the study, MACS gave an optimum length of peaks and a good number (much more than QuEST) of peaks. The higher number makes it easier to select the peaks based on multiple parameters, while a lower number of peaks (even if they have better reproducibility) leaves the experimenter with little options. The fact that MACS uses a bimodal distribution of tags, which essentially would report binding sites only and also the fact that it could be used both with a control and without a control sample made it an obvious choice

for peak finding algorithm. Finally out of the three datasets that were used for validation by qPCR,all the peaks MACS, were most enriched. Hence MACS was used for peak calling in the study.

Materials & methods

3.2: Quality control for the sequences

FastQC is a program that can be used as a tool for quality control of high throughput sequencing data. The program was used to check the integrity and quality of the short 36 base pair sequences obtained from Illumina Sequencer.

Per base sequence quality shows the range of quality values across all bases at each position in the sequence file. A high quality value represents a better base call. Usually towards the end of the read the quality score goes down (especially the last base).

Per sequence quality score tells what subset of the reads have universally low quality reads. Only a small percentage of the reads should have a low quality.

Sequence length distribution tells about the fraction of reads that have a particular length. Usually all the sequence reads have more or less uniform length but in some cases, they may be erratic indicating either fault in library preparation or in sequencing.

Per base GC content shows the GC content of each base position in the sequence file. In a random library one would not expect to have a huge difference between bases at any position in the sequence file. The graph therefore should come out more or less horizontal to the X- axis with the GC content underlying the GC content of the genome. A change in GC bias in independent bases may indicate over represented sequences in the sequence run.

The module *sequence duplication levels* shows the relative number of sequences with different degrees of duplication. A low level of duplication shows a high level of coverage of the target sequence, but a high level of duplication is more likely to

indicate some kind of enrichment bias (PCR over amplification or any other artefacts).

The module (*over represented kmers*) counts the enrichment of every 5 mer within the sequence file. After calculating an expected frequency at which this kmer should have been seen based on the GC content, it uses the actual count to represent an observed/ expected ratio of the mer. Kmers show a general bias or enrichment in the sequence.

Per base sequence content plots out the proportion of each base position in a file for which each of the four normal DNA bases have been called. The line of each of the four bases usually run parallel to each other as in a normal DNA sequence library there shouldn't be any difference between the different bases of a sequence run. The relative amount of each base also represents the GC content of the sample.

Per sequence GC content measures the GC content across the whole file and compares it to a normal modelled distribution of GC plot. An unusually shifted library indicates a contaminated library or some other kinds of biased subset.

Per base N content essentially tells us percentage of base calls at each position for which N was called. N is called by a sequencer when it is unable to determine accurately the base at that position. A low proportion of N is fine across the sequence run but a higher proportion of N indicates fault in sequencing.

3.3: Indexing the genome

Bowtie- build indexer was used to make an index of the *Apis mellifera* (version 2.0) genome. In order to search for matches between short sequences (36 mers in this case) and sequence database comprising full genomes, the full genome is usually indexed in order to provide faster searches for matches. The index file is later needed to align the short reads to the genome.

3.4: Alignment of the short reads to the genome

Bowtie was used to align the short reads to the genome. It is a fast, memory – efficient short read aligner that maps the short reads to the genome. Bowtie works best while aligning short sequences to the whole genome and can take sequences of length 1024 bases for alignment.

Bowtie looks for alignments, which do not have more than two mismatches in the first 28 bases on the high quality end of the read. The first 28 bases are called the "seed". It further imposes a cut off based on Phred quality. The sum of all mismatched positions across all 36 bases shouldn't exceed 70, else the alignment is not reported. This is very obvious as bowtie excludes with the help of this option any high confidence mismatches and keeps only the mismatches, which might show some ambiguity in the base score. If there are many alignments matching the criteria, bowtie reports the one which is better in quality based on the above two criteria. Bowtie was also used to report one random alignment when multiple alignments for a read were possible. Bowtie gave the necessary alignments in SAM format.

SAM format is a generic format for storing large nucleotide sequence alignments. It allows the file to be indexed by genomic position to efficiently retrieve all reads aligning to a locus. The SAM file is converted to BAM format and then sorted and indexed for fast random access.

3.5: Peak detection

Model-based analysis of ChIP-Seq (MACS) version 1.4 was used to find peaks from the aligned sequences. The ChIP-DNA fragments are sequenced from both ends and hence the tag density around a true binding site should always show a bimodal distribution pattern. The Watson tags should be enriched on one side of the binding region whereas the crick tags should be enriched on the other side of the binding region. MACS takes advantage of this bimodal distribution to precisely locate the binding site. MACS takes the sonication size as input and finds regions in the

genome with tags more than a particular enrichment (defined by mfold) relative to a random tag genome distribution. It then uses 1000 of these peaks to make a model, separates their Watson and Crick tags and determines the distance between their centres. This distance is referred to as ‘d’.

Next for peak detection, MACS scales the control tag count towards the experimental tag count. MACS also subsequently removes duplicate tags which might arise during biased PCR amplification. Now, the program shifts every tag by d/2, and then MACS slides 2d window across the whole genome to find peaks above significant tag enrichment. Overlapping peaks are merged and each tag position is extended to d distance from its center. The distribution of tags in the control sample and experimental sample are used to find the fold enrichment and FDR for each peak. Workflow of MACS has been outlined in **plate 3.19**.

3.6: Selecting Peaks

Plate 3.28 represents diagrammatically the way in which peaks from the two replicates were selected. First, pre immune IgG Peaks were subtracted from the anti Ubx IgG peaks. Of the remaining peaks only those peaks were selected which had an FDR <= 1% and fold enrichment over input greater than 10. This process was carried out for both the replicates. Amongst the filtered peaks only the peaks that were common to both the replicates were used for subsequent applications. Bedtools was used for various operations on genomic intervals.

Results & Discussion

3.7: Quality control for the sequences (Summarised in plates 3.7-3.18)

FASTQC was used to check the quality of sequences as mentioned in materials and methods. A quick summary of the results obtained for various reads has been mentioned below:

a. Per base sequence quality:

Per base sequence quality of all the 36 bases for the reads obtained from various samples of the two replicates had a high score except the last base. This indicated good quality sequencing for most of the 36 bases and hence the reads could be used for further analysis.

b. Quality score distribution across all the sequences indicated high quality reads for most of the sequences for all the samples.

c. Sequence length of most of the sequences was around 36 bases which once again indicated good sequencing quality.

d. GC content across all the bases in all the sequence runs remained fairly constant and showed a GC value of 40 % indicating the high AT content of the Apis genome, which is consistent with the genome sequence of *Apis mellifera*. The graph also didn't show much spikes for any of the replicates suggestive of lack of contamination in the library.

e. Sequence duplication level was low for all the input samples but was pretty high for the chipped DNA samples. This could be due to the lower amount of DNA in chipped samples and also due to PCR bias. However, as the subsequent bioinformatics tools remove duplicate tags and also take care of the difference in size of datasets, the sequences were used for further analysis

f. kmer content varied from sample to sample but was ignored as the over representation of these mers would ultimately be removed by the downstream processing.

g. Sequence content across all bases once again remained fairly constant and indicated AT rich genome and didn't show spikes indicative or few or little contaminating DNA sequences.

h. GC distribution across all the bases too followed almost a normal distribution pattern for all the samples.

i. N content across all the bases was negligible which once again indicated good sequencing.

3.8: Indexing the genome

The genome was indexed using Bowtie-build after removing empty sequences and the output files were used for subsequent alignment.

3.9: Alignment of the short reads to the indexed genome

Bowtie was used to align the short reads to the genome and the output file was stored as SAM format, which was subsequently converted to BAM format. The BAM file was subsequently sorted and indexed. The reads that aligned to the genome and those that failed to align have been summarised in **table 1 and 2.**

3.10: Peak detection and selection

MACS was used to detect peaks in the aligned sequences. The peak models obtained from various alignments have been described in **plate**s 3.20- 3.27. In all the cases, input DNA of the respective sample was taken as a control and enrichment and False discovery rate (FDR) were calculated using the input. The selection of peaks has already been described in the earlier section. A graphical representation of the peaks

obtained in two replicates of ChIP has been shown in **plate** 3.29; while the peaks common to two replicates have been summarised in **plate** 3.30.

Finally, a cut off of 10 fold enrichment for peaks and FDR<= 1 % was applied and 2350 peaks were found common to both replicates of hind wing at this high confidence level, while 927 peaks were found common to both replicates of fore wing. 526 of these peaks (the high confidence peaks of fore and hind wing) were common between the fore and hind wing.

Summary

To summarise, this chapter has described the various means by which the quality of the short sequences have been analysed, their alignment to the honey bee genome and detection of peaks. Overall, 2250 peaks (regions bound by Ubx) were reported in the hind wing discs of *Apis mellifera* while 927 peaks were detected in the fore wing discs of *Apis mellifera* amongst which 526 were shared between the two datasets.

In the next chapter, the genes associated with these peaks (targets of Ubx in honey bee) have been described and compared with those in the fruit fly.

Plates

Chapter 3

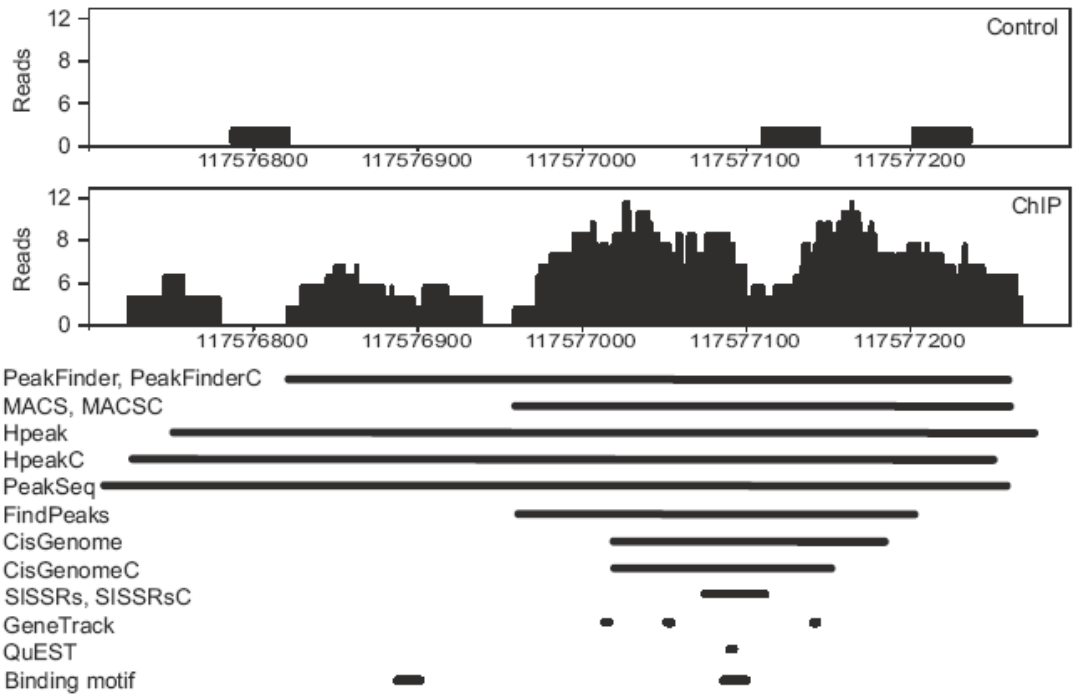


Plate 3.1: Characteristics of detected binding region by various algorithms (the data is for STAT6 transcription factor, adapted from Laajala *et. al.*, 2009). The number of overlapping reads for either the control of ChIP experiment is shown on the Y axis while the genomic co- ordinate is shown on the X axis. Smallest peaks are identified by GeneTrack and QuEST while longest are identified by PeakSeq.

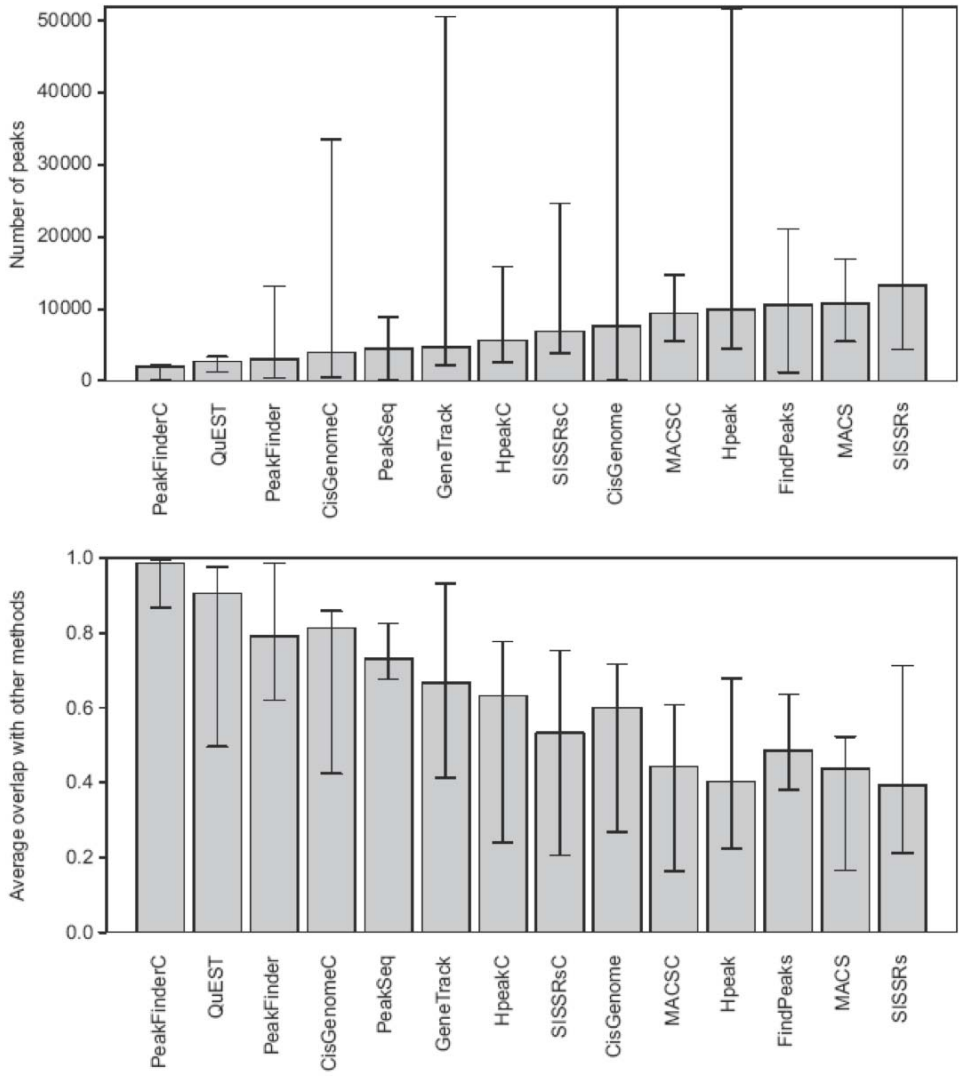


Plate 3.2: Number and overlap of detected peaks (Courtesy: Laajala *et. al.*, 2009). Figure on top shows the average number of peaks detected by each algorithm with maximum and minimum number indicated by error bars. The lower figure indicates the overlap of peaks identified by the particular algorithm with those found by other algorithms.

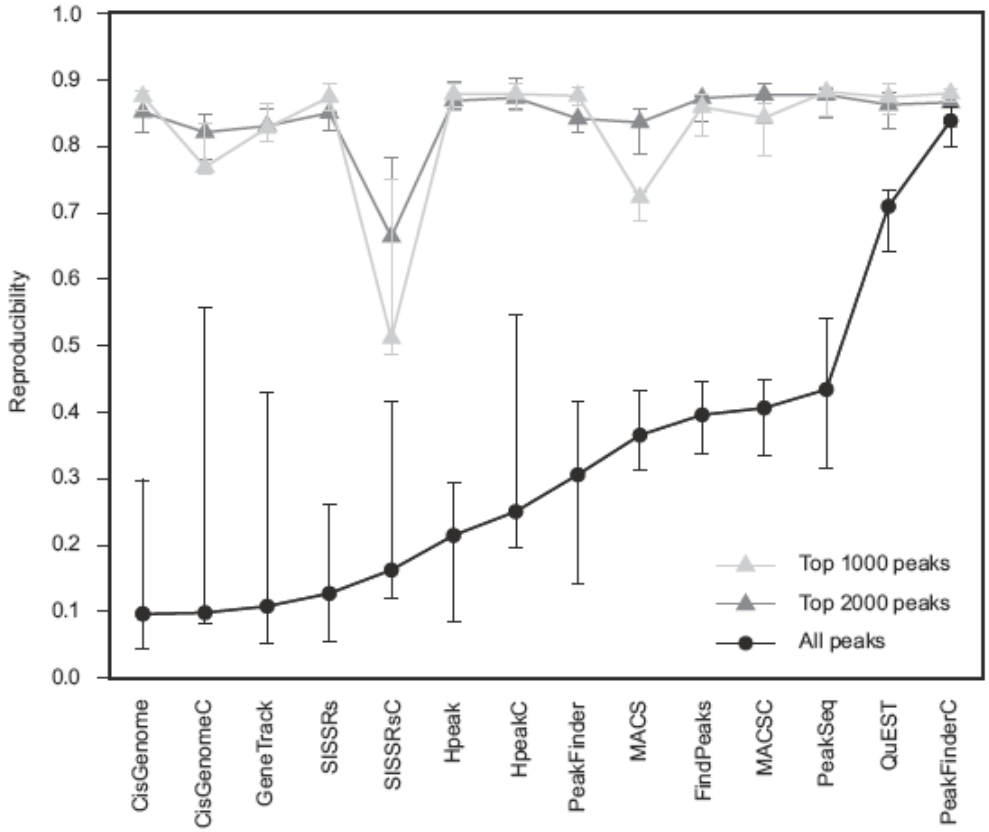


Plate 3.3: Reproducibility of detections across three NSRF samples (adapted from Laajala *et. al.*, 2009).

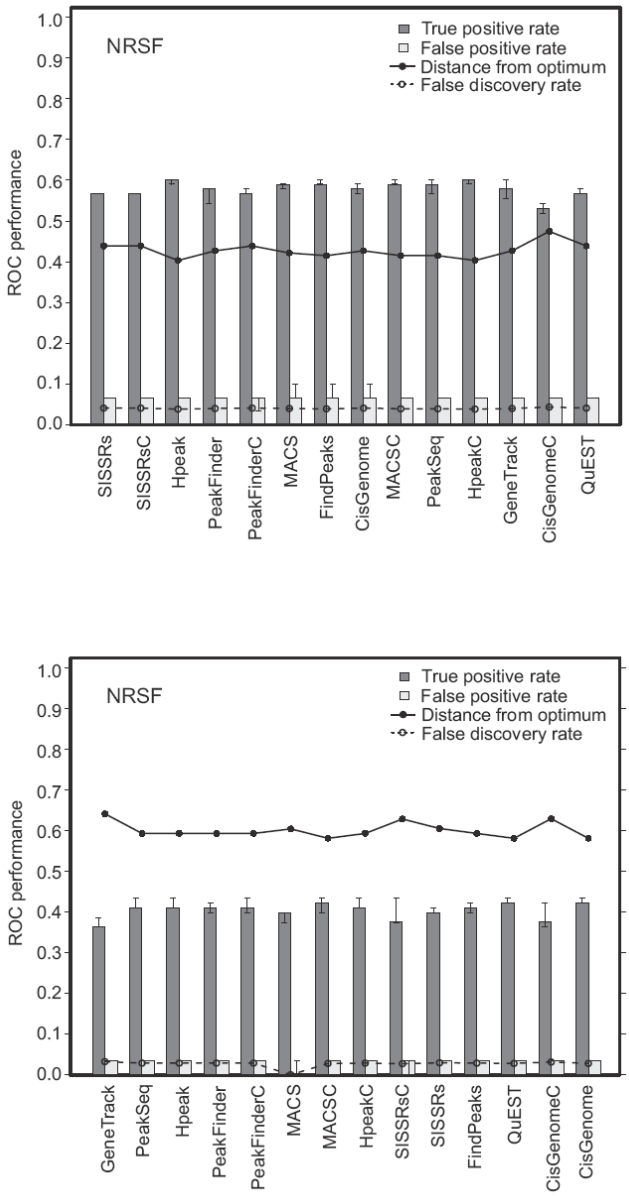


Plate 3.4: External validation of the predicted binding sites using qPCR (data for NSRF binding sites is shown in the image, the top panel shows performance while using all the peaks while the bottom panel indicates the performance when only top 1000 peaks are considered). True positive rates are shown by dark bars while false positive rate has been shown with light grey bars. The solid lines indicate the distance from optimal performance while the dashed line indicates FDR.

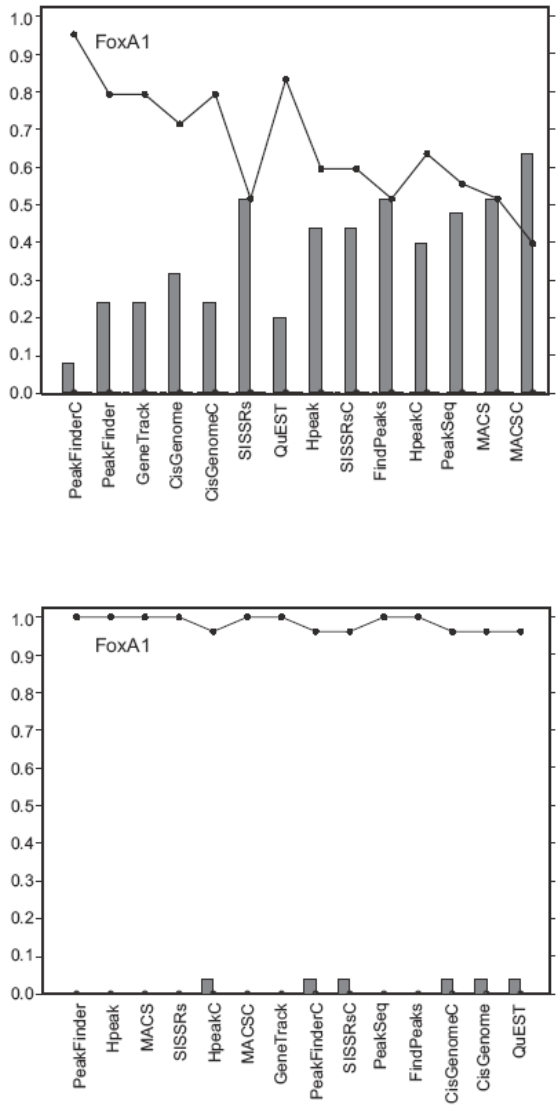


Plate 3.5: External validation of the predicted binding sites using qPCR (data for FoxA1 binding sites is shown in the image, the top panel shows performance while using all the peaks while the bottom panel indicates the performance when only top 1000 peaks are considered). True positive rates are shown by dark bars while false positive rate has been shown with light grey bars. The solid lines indicate the distance from optimal performance while the dashed line indicates FDR.

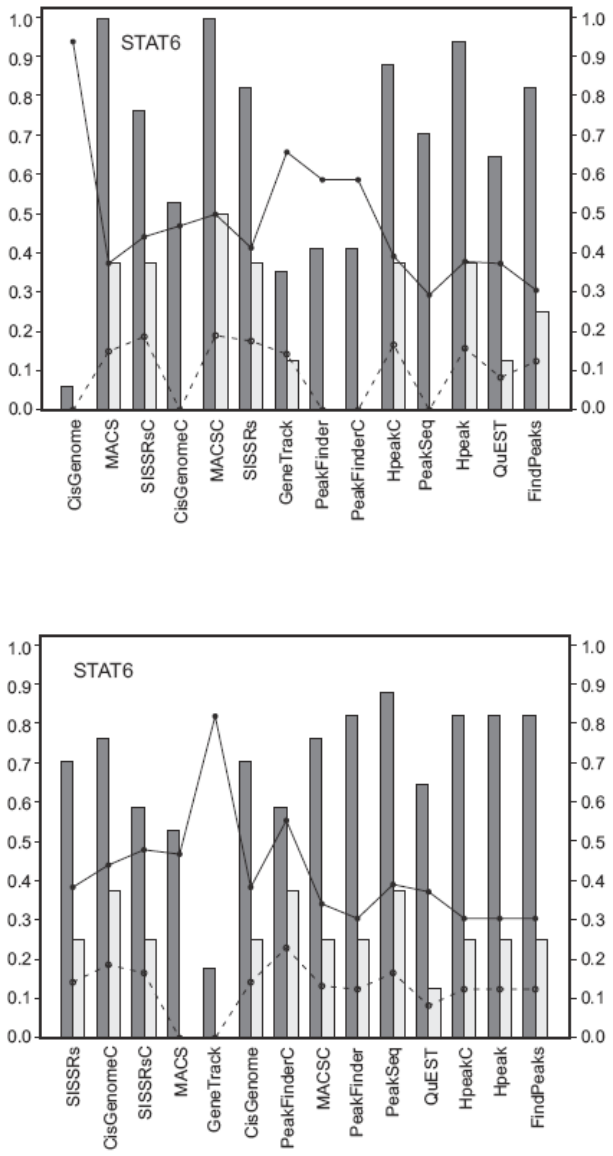


Plate 3.6: External validation of the predicted binding sites using qPCR (data for STAT6 binding sites is shown in the image, the top panel shows performance while using all the peaks while the bottom panel indicates the performance when only top 1000 peaks are considered). True positive rates are shown by dark bars while false positive rate has been shown with light grey bars. The solid lines indicate the distance from optimal performance while the dashed line indicates FDR.

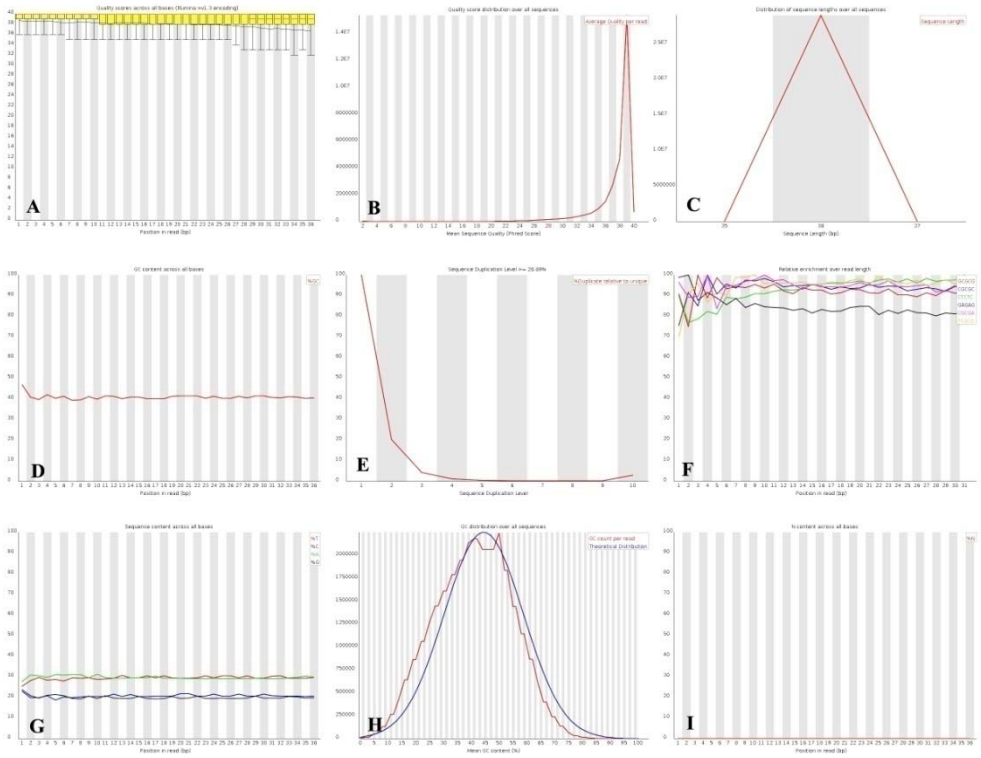


Plate 3.7: FASTQC output of forewing 10 input DNA reads

A. Per base sequence quality

B. Per sequence quality scores.

C. Sequence length distribution

D. Per base GC content

E. Sequence duplication levels

F. Kmer content

G. Per base sequence content

H. Per sequence GC content

I. Per base N content

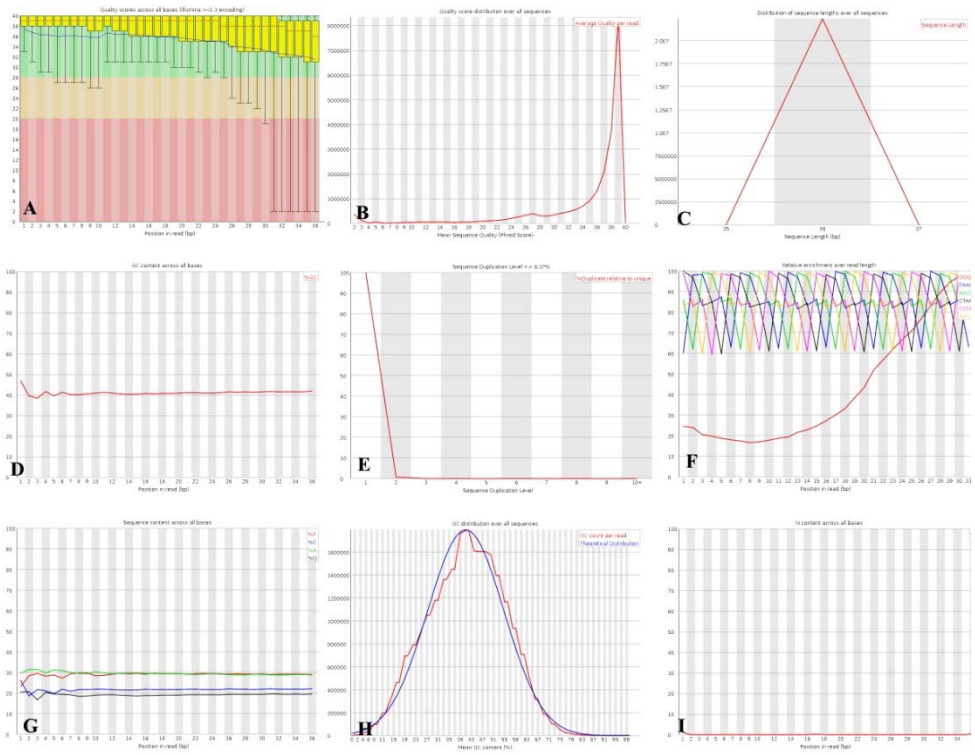


Plate 3.8: FASTQC output of forewing 8 input DNA reads

A. Per base sequence quality

B. Per sequence quality scores.

C. Sequence length distribution

D. Per base GC content

E. Sequence duplication levels

F. Kmer content

G. Per base sequence content

H. Per sequence GC content

I. Per base N content

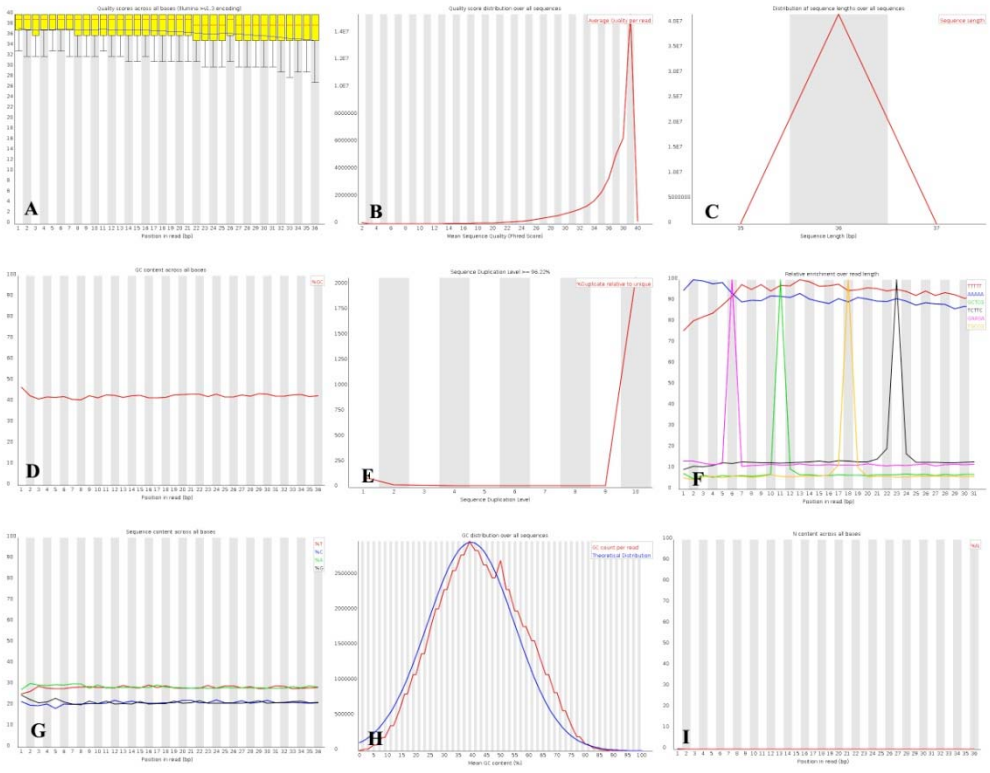


Plate 3.9: FASTQC output of forewing 10 anti Ubx chipped DNA reads

A. Per base sequence quality

B. Per sequence quality scores.

C. Sequence length distribution

D. Per base GC content

E. Sequence duplication levels

F. Kmer content

G. Per base sequence content

H. Per sequence GC content

I. Per base N content

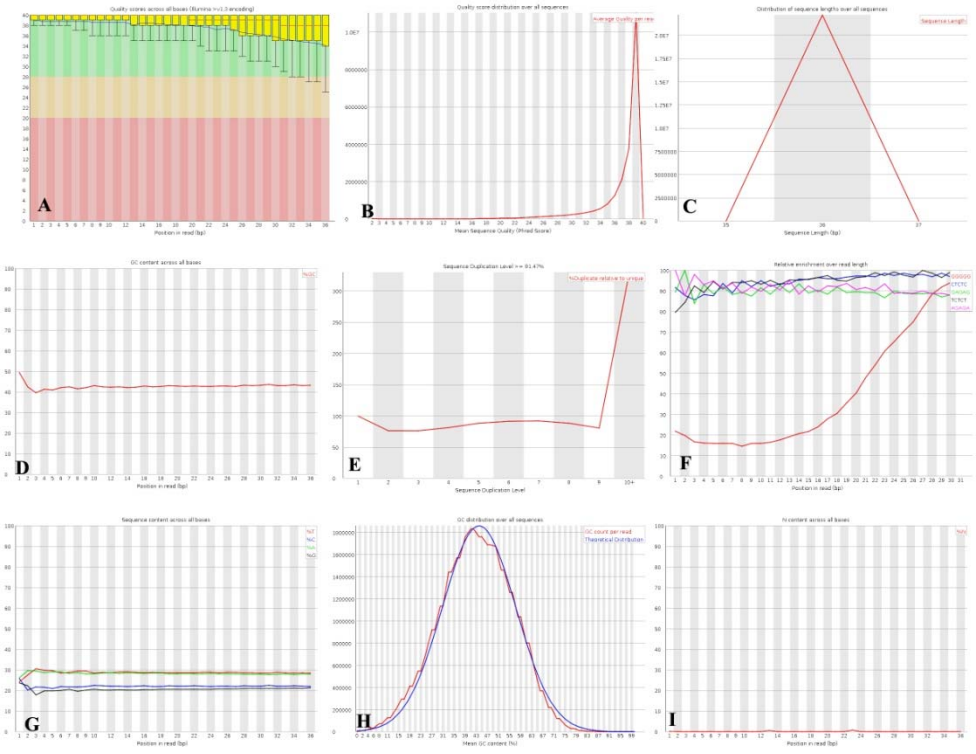


Plate 3.10: FASTQC output of forewing 8 anti Ubx chipped DNA reads

A. Per base sequence quality

B. Per sequence quality scores.

C. Sequence length distribution

D. Per base GC content

E. Sequence duplication levels

F. Kmer content

G. Per base sequence content

H. Per sequence GC content

I. Per base N content

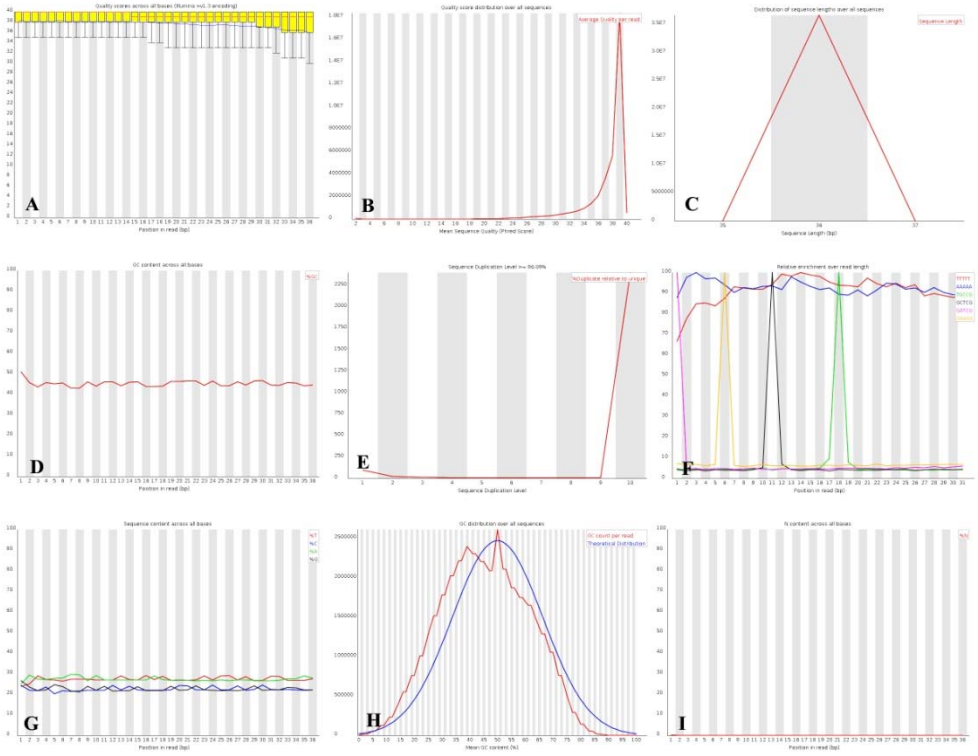


Plate 3.11: FASTQC output of forewing 10 pre immune IgG pull down DNA reads

A. Per base sequence quality

B. Per sequence quality scores.

C. Sequence length distribution

D. Per base GC content

E. Sequence duplication levels

F. Kmer content

G. Per base sequence content

H. Per sequence GC content

I. Per base N content

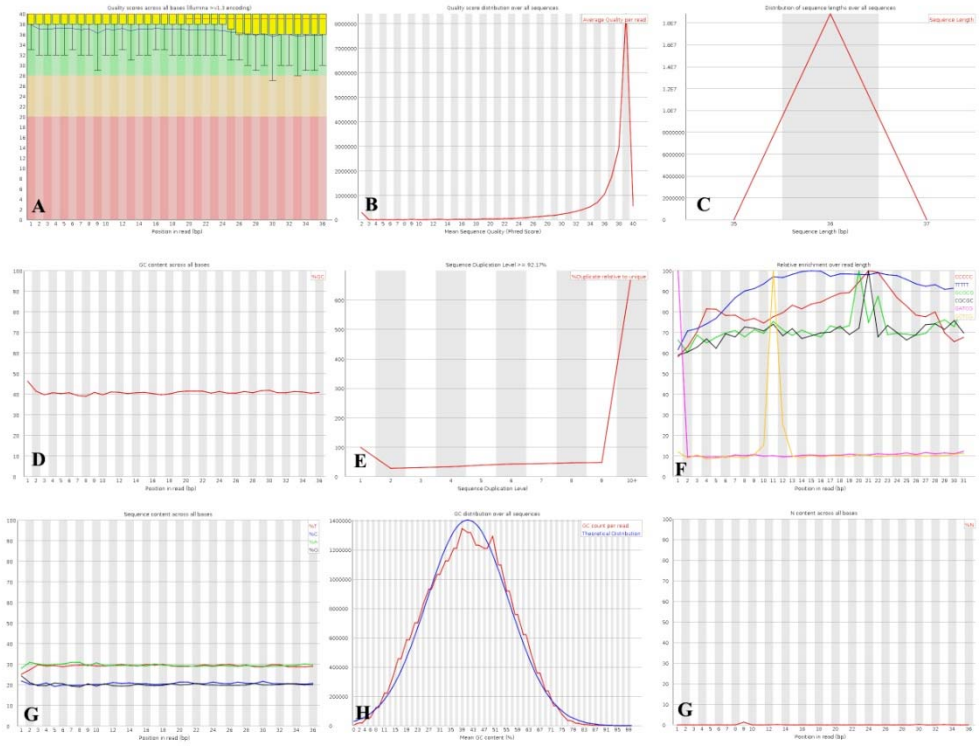


Plate 3.12: FASTQC output of forewing 8 pre immune IgG pull down DNA reads

A. Per base sequence quality

B. Per sequence quality scores.

C. Sequence length distribution

D. Per base GC content

E. Sequence duplication levels

F. Kmer content

G. Per base sequence content

H. Per sequence GC content

I. Per base N content

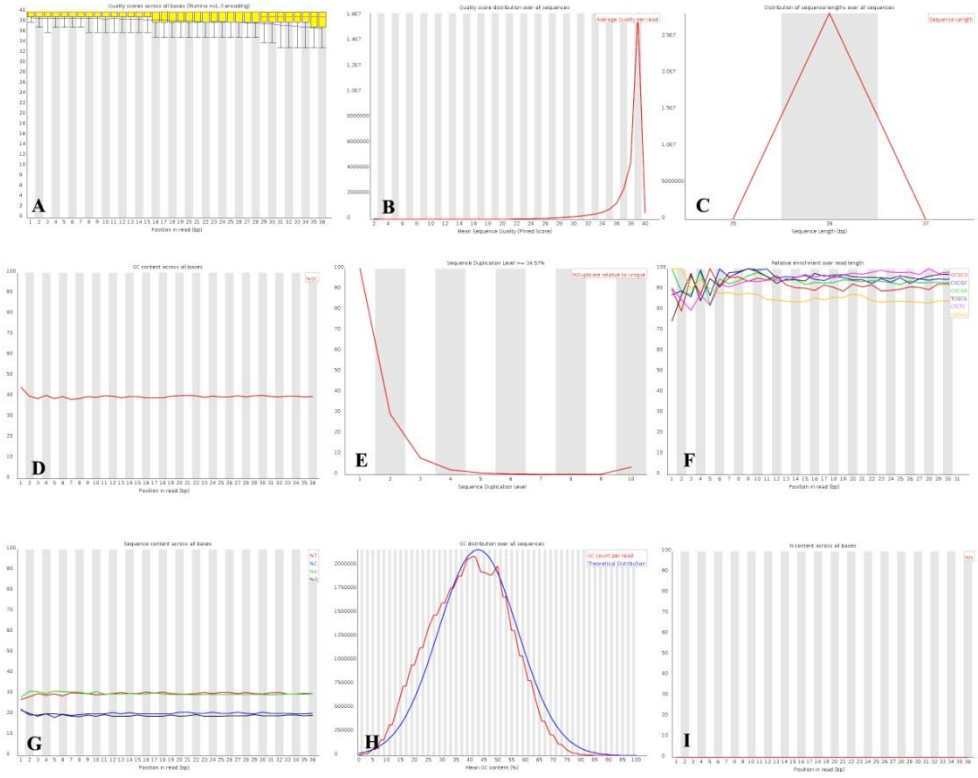


Plate 3.13: FASTQC output of hind wing 10 input DNA reads

A. Per base sequence quality

B. Per sequence quality scores.

C. Sequence length distribution

D. Per base GC content

E. Sequence duplication levels

F. Kmer content

G. Per base sequence content

H. Per sequence GC content

I. Per base N content

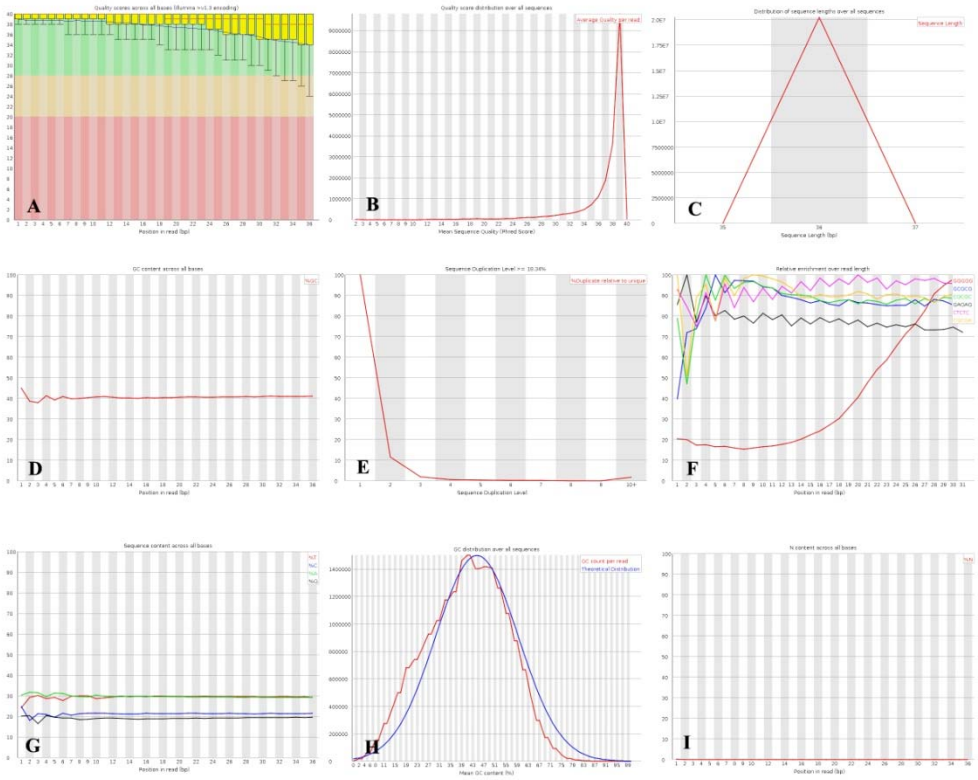


Plate 3.14: FASTQC output of hind wing 8 input DNA reads

A. Per base sequence quality

B. Per sequence quality scores.

C. Sequence length distribution

D. Per base GC content

E. Sequence duplication levels

F. Kmer content

G. Per base sequence content

H. Per sequence GC content

I. Per base N content

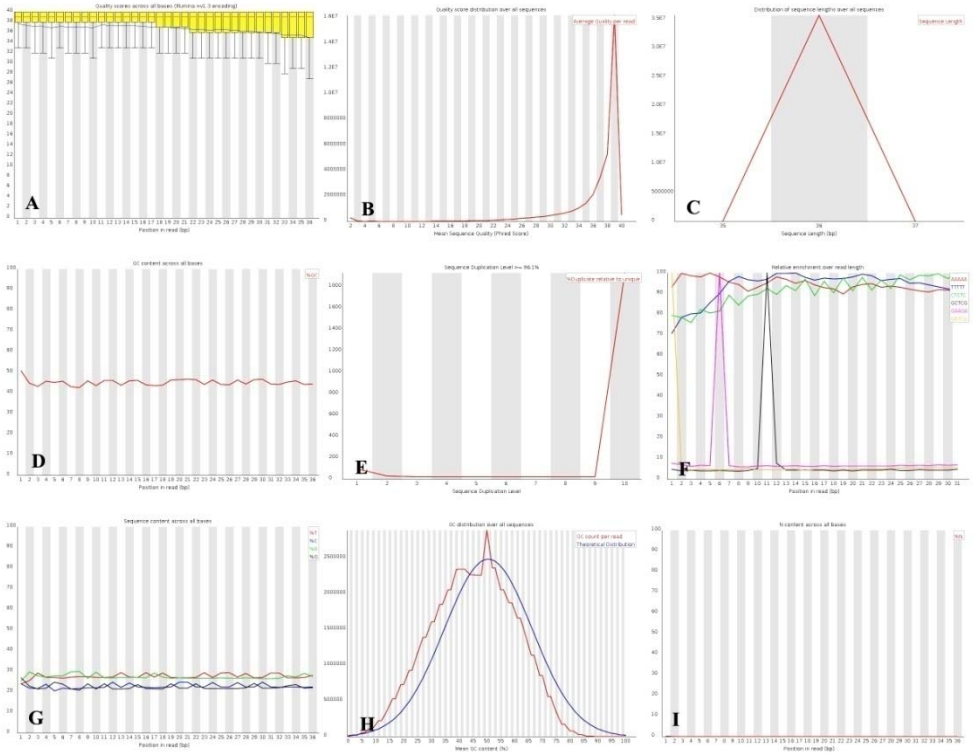


Plate 3.15: FASTQC output of hind wing 10 anti Ubx chipped DNA reads

A. Per base sequence quality

B. Per sequence quality scores.

C. Sequence length distribution

D. Per base GC content

E. Sequence duplication levels

F. Kmer content

G. Per base sequence content

H. Per sequence GC content

I. Per base N content

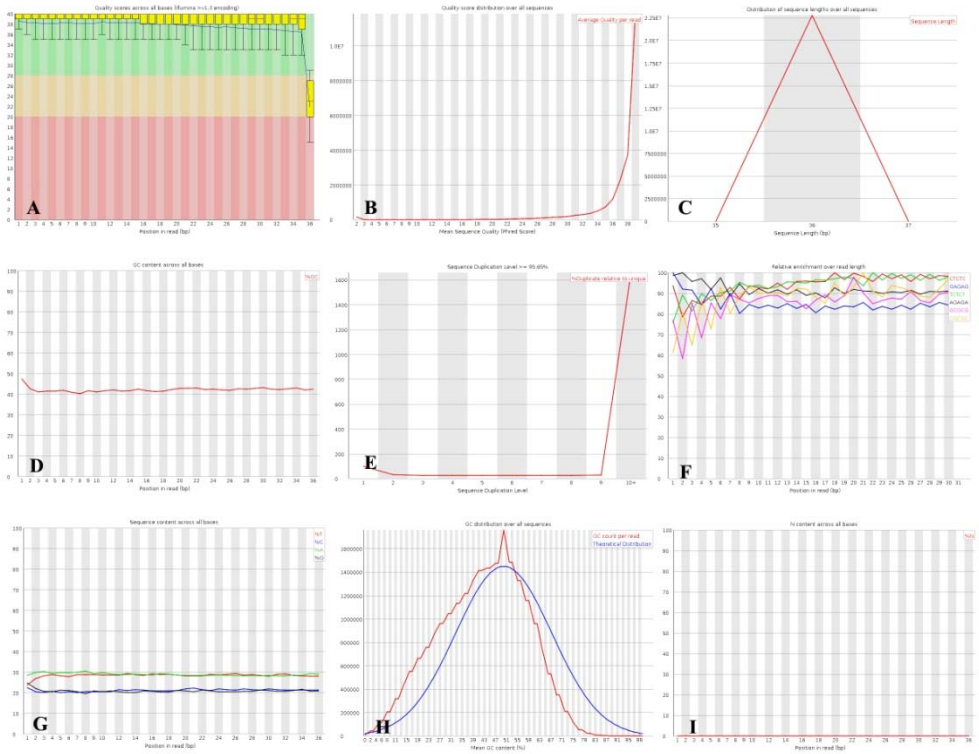


Plate 3.16: FASTQC output of hind wing 8 anti Ubx chipped DNA reads

A. Per base sequence quality

B. Per sequence quality scores.

C. Sequence length distribution

D. Per base GC content

E. Sequence duplication levels

F. Kmer content

G. Per base sequence content

H. Per sequence GC content

I. Per base N content

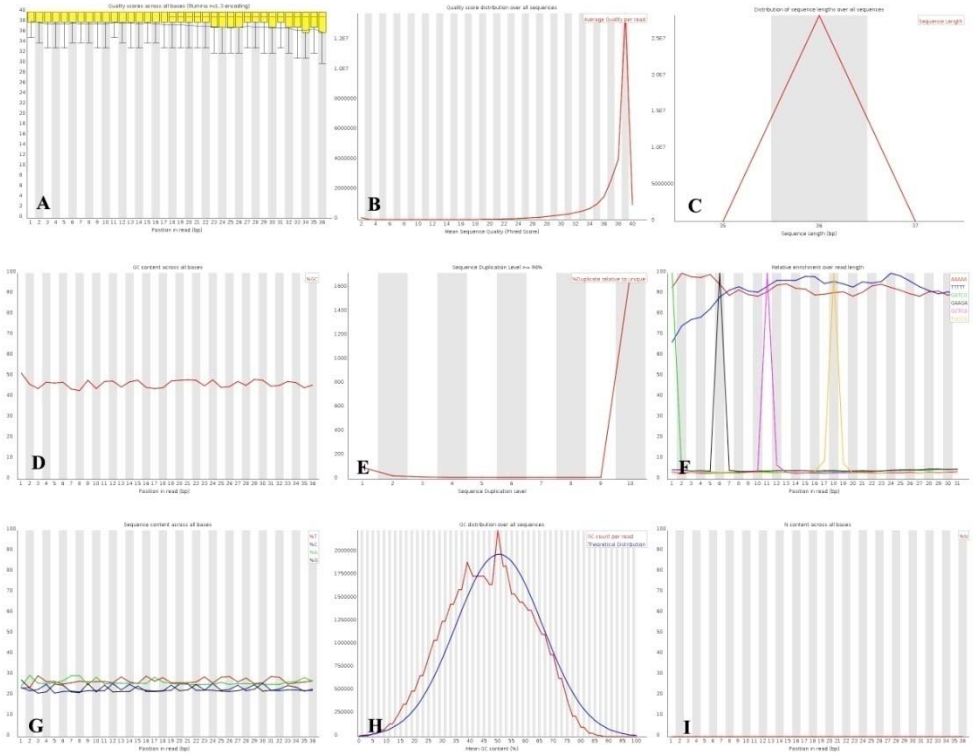


Plate 3.17: FASTQC output of hind wing 10 pre immune IgG pull down DNA reads

- **A.** Per base sequence quality
- **B.** Per sequence quality scores.
- **C.** Sequence length distribution
- **D.** Per base GC content
- **E.** Sequence duplication levels
- **F.** Kmer content
- **G.** Per base sequence content
- **H.** Per sequence GC content
- **I.** Per base N content

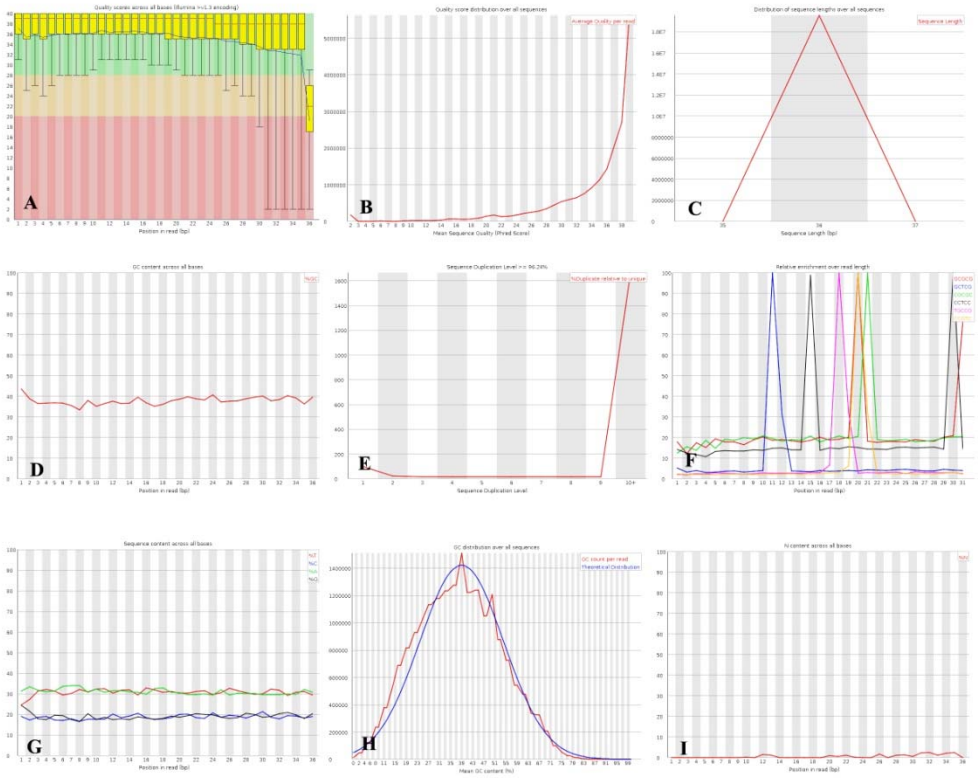


Plate 3.18: FASTQC output of hind wing 8 pre immune IgG pull down DNA reads

A. Per base sequence quality

B. Per sequence quality scores.

C. Sequence length distribution

D. Per base GC content

E. Sequence duplication levels

F. Kmer content

G. Per base sequence content

H. Per sequence GC content

I. Per base N content

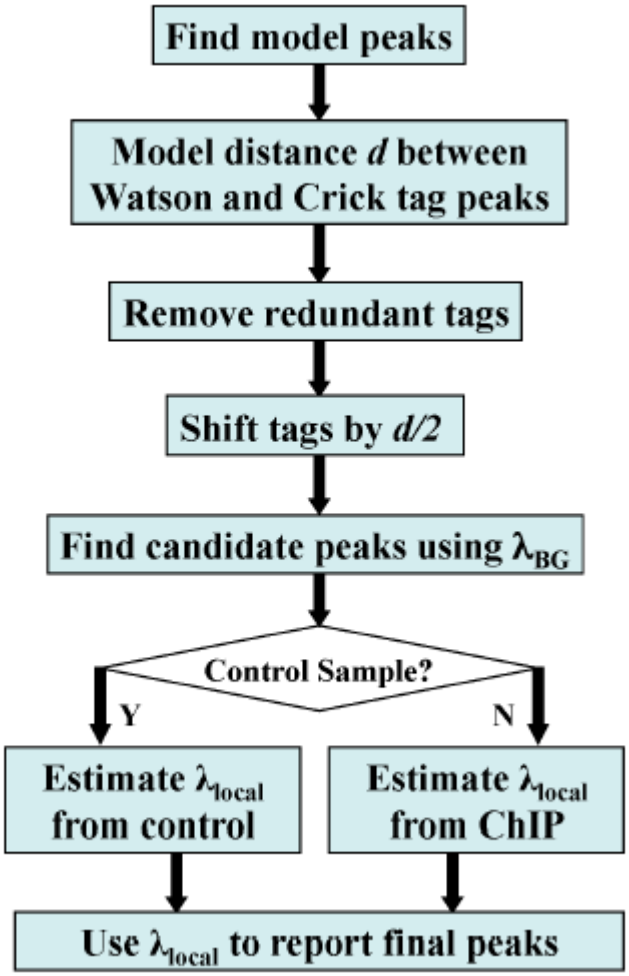


Plate 3.19: Workflow used by MACS to report peaks

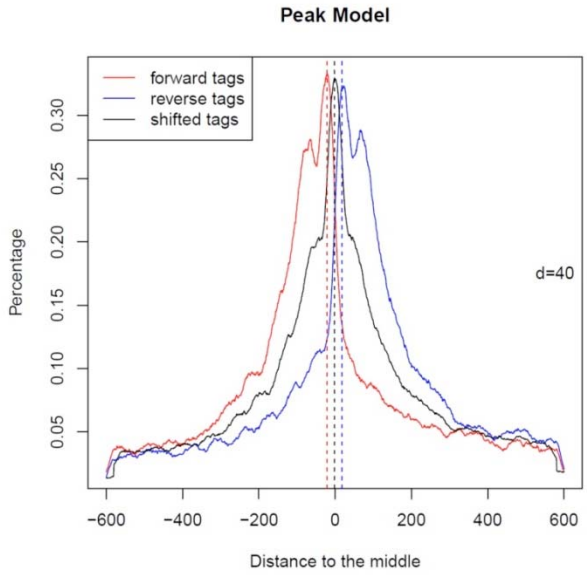


Plate 3.20: Peak model obtained from MACS for fore wing 10 anti Ubx chipped DNA (compared to input)

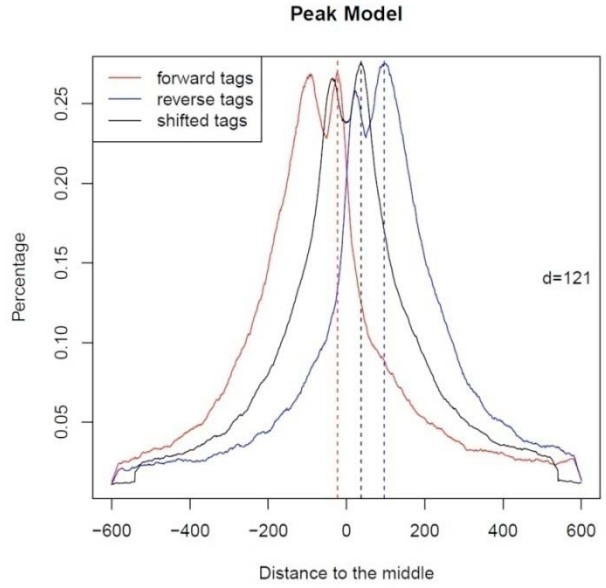


Plate 3.21: Peak model obtained from MACS for fore wing 8 anti Ubx chipped DNA (compared to input)

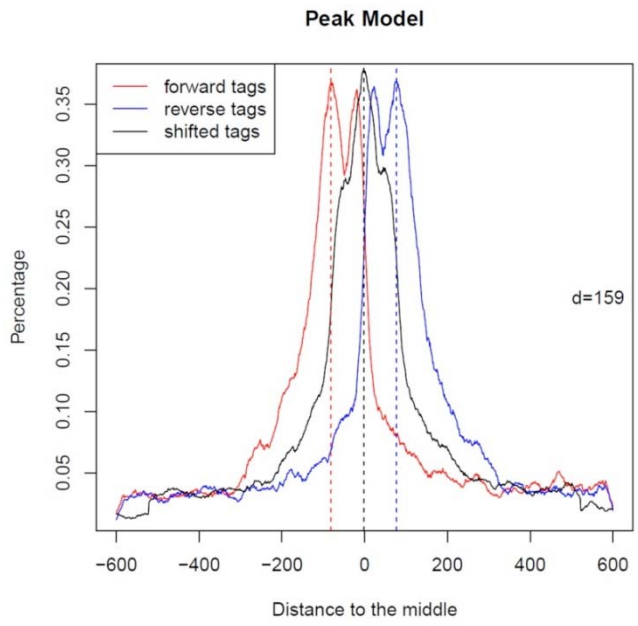


Plate 3.22: Peak model obtained from MACS for fore wing 10 pre immune IgG chipped DNA (compared to input)

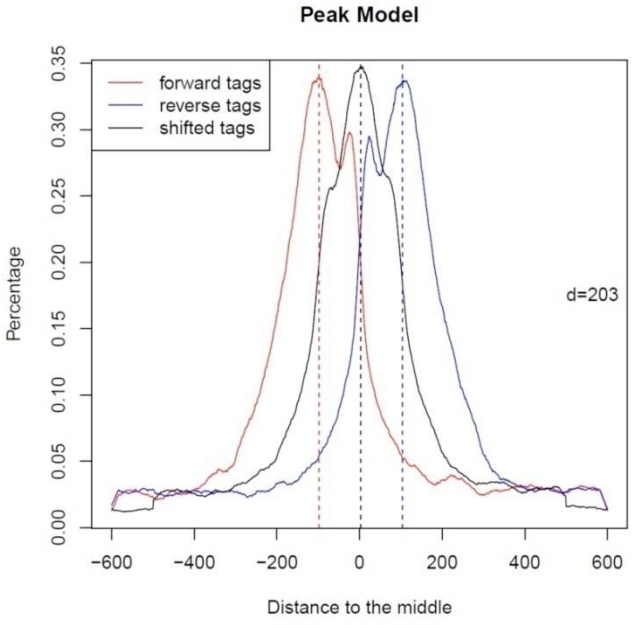


Plate 3.23: Peak model obtained from MACS for fore wing 8 pre immune IgG chipped DNA (compared to input)

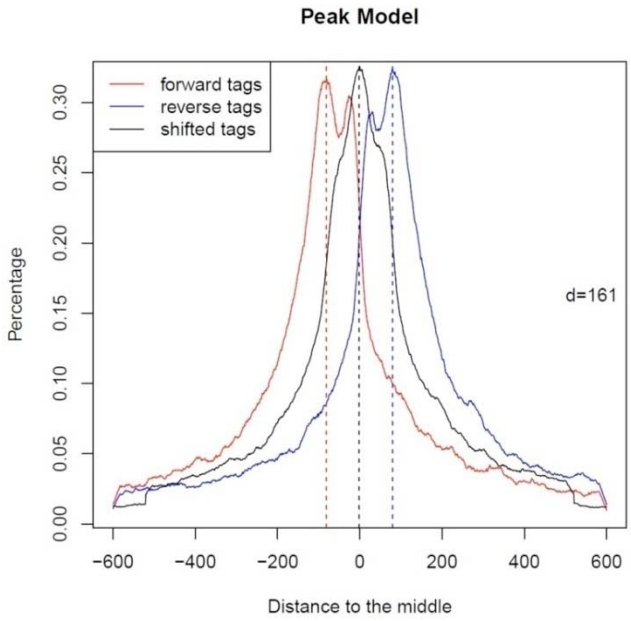


Plate 3.24: Peak model obtained from MACS for hind wing 10 anti Ubx chipped DNA (compared to input)

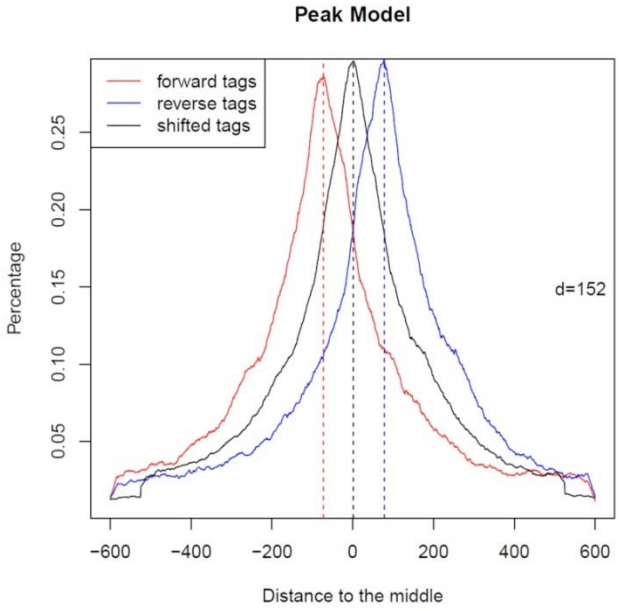


Plate 3.25: Peak model obtained from MACS for hind wing 10 anti Ubx chipped DNA (compared to input).

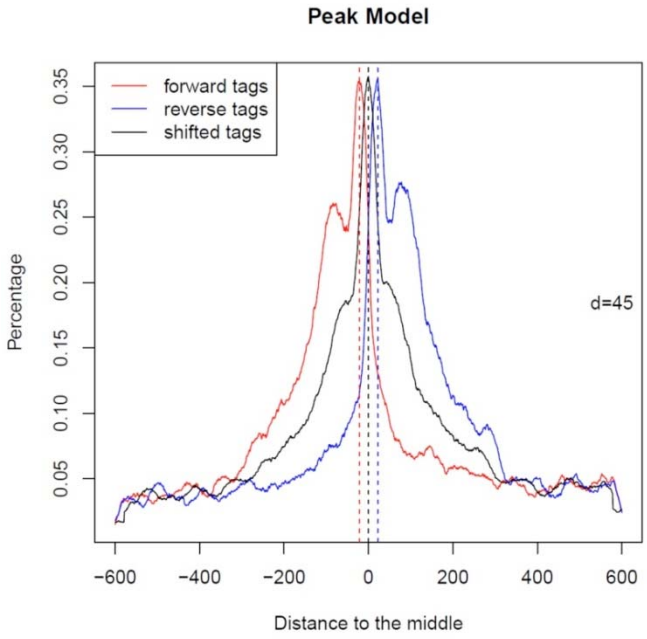


Plate 3.26: Peak model obtained from MACS for hind wing 10 pre immune IgG chipped DNA (compared to input)

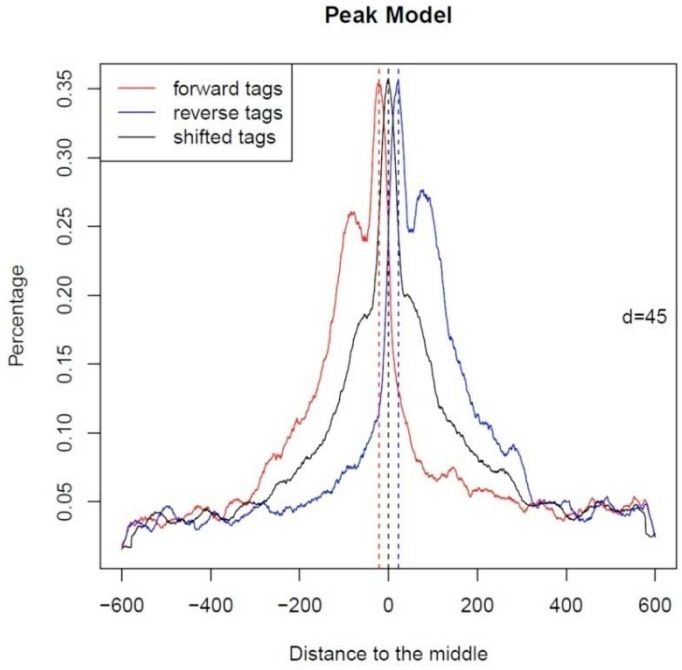
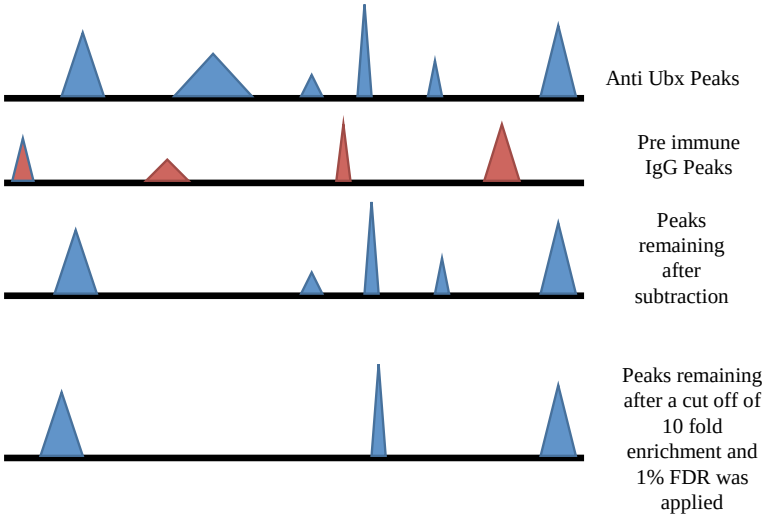
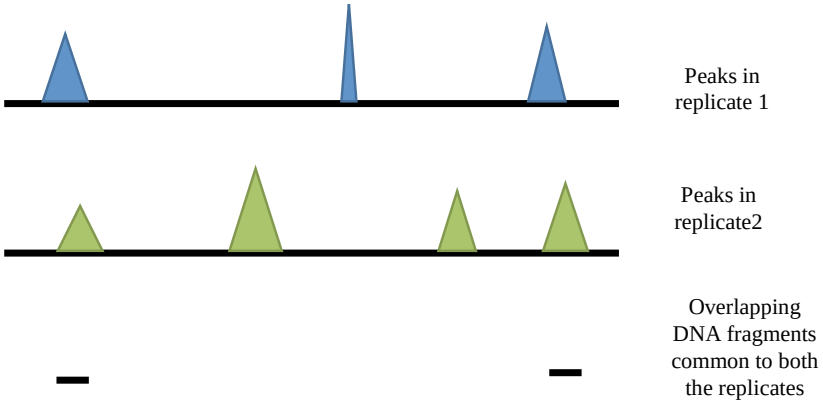


Plate 3.27: Peak model obtained from MACS for fore wing 8 pre immune IgG chipped DNA (compared to input).



Filtering Peaks



Selecting common peaks

Plate 3.28: Final Peaks

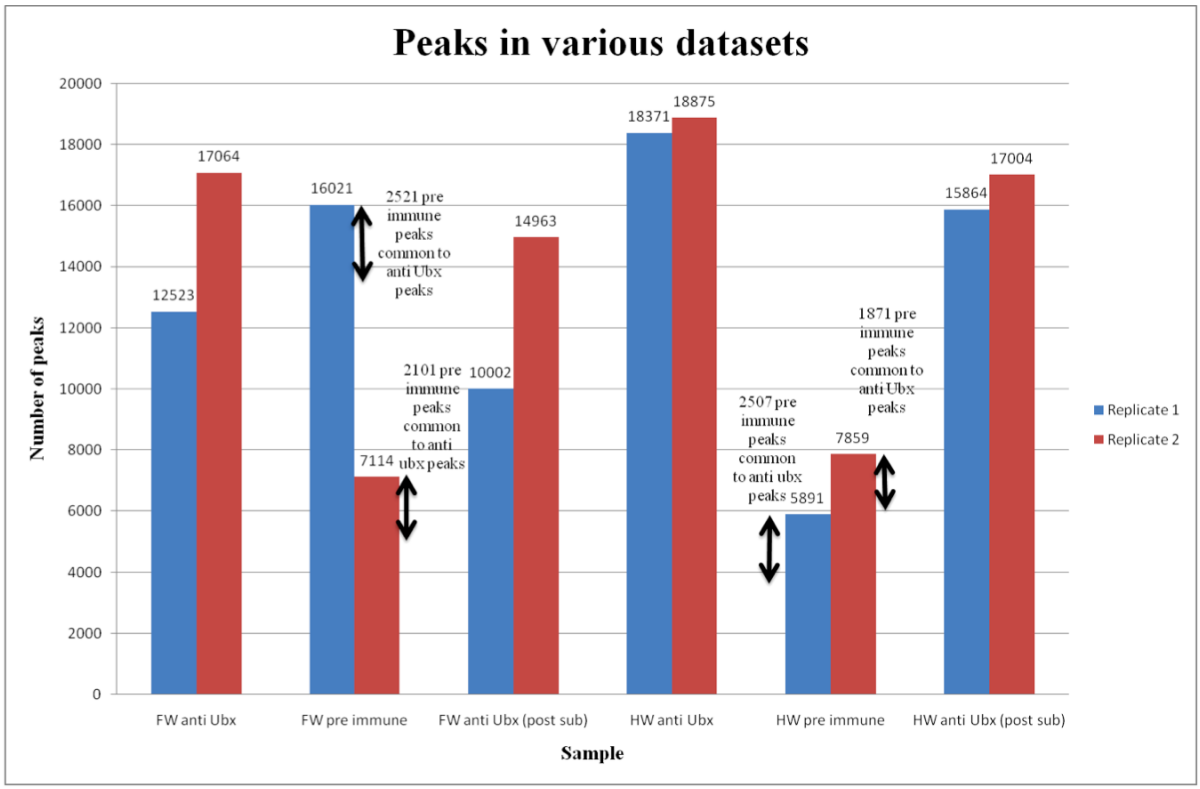


Plate 3.29: Graph showing peaks obtained in various datasets

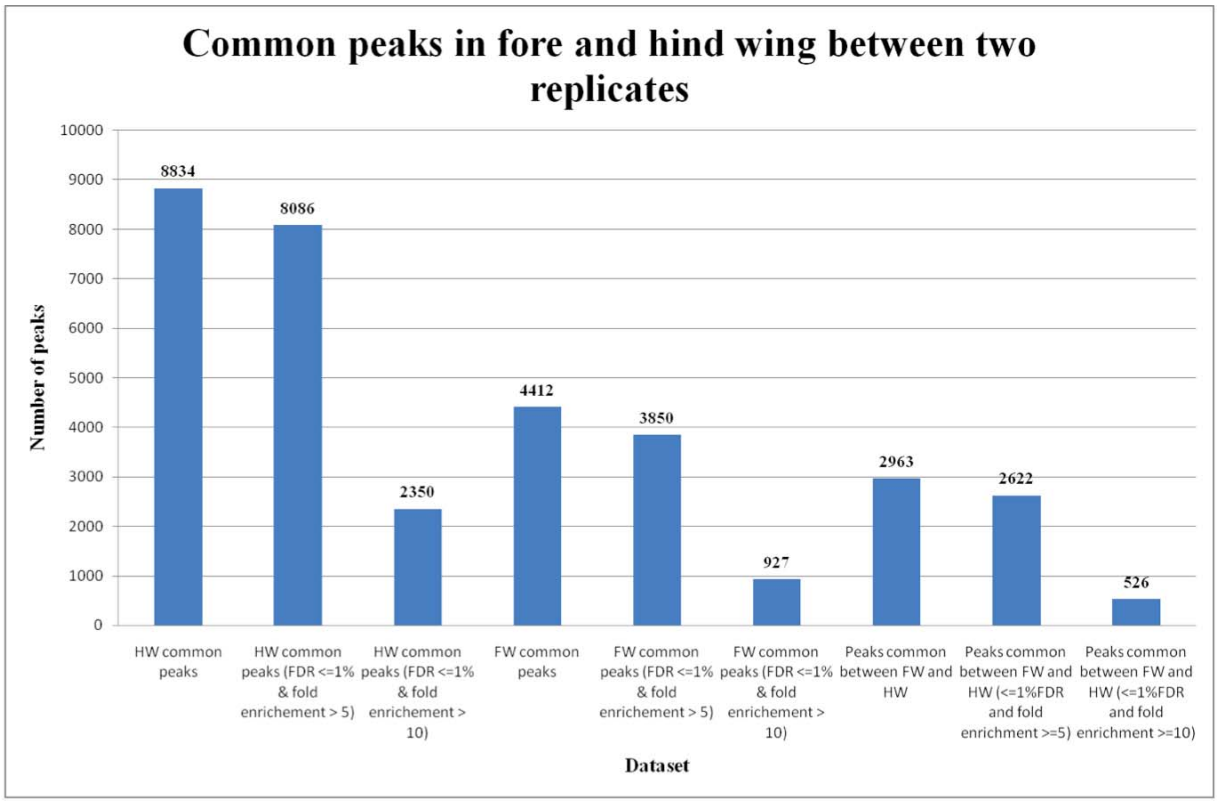


Plate 3.30: Graph showing common peaks in fore and hind wing for two replicates

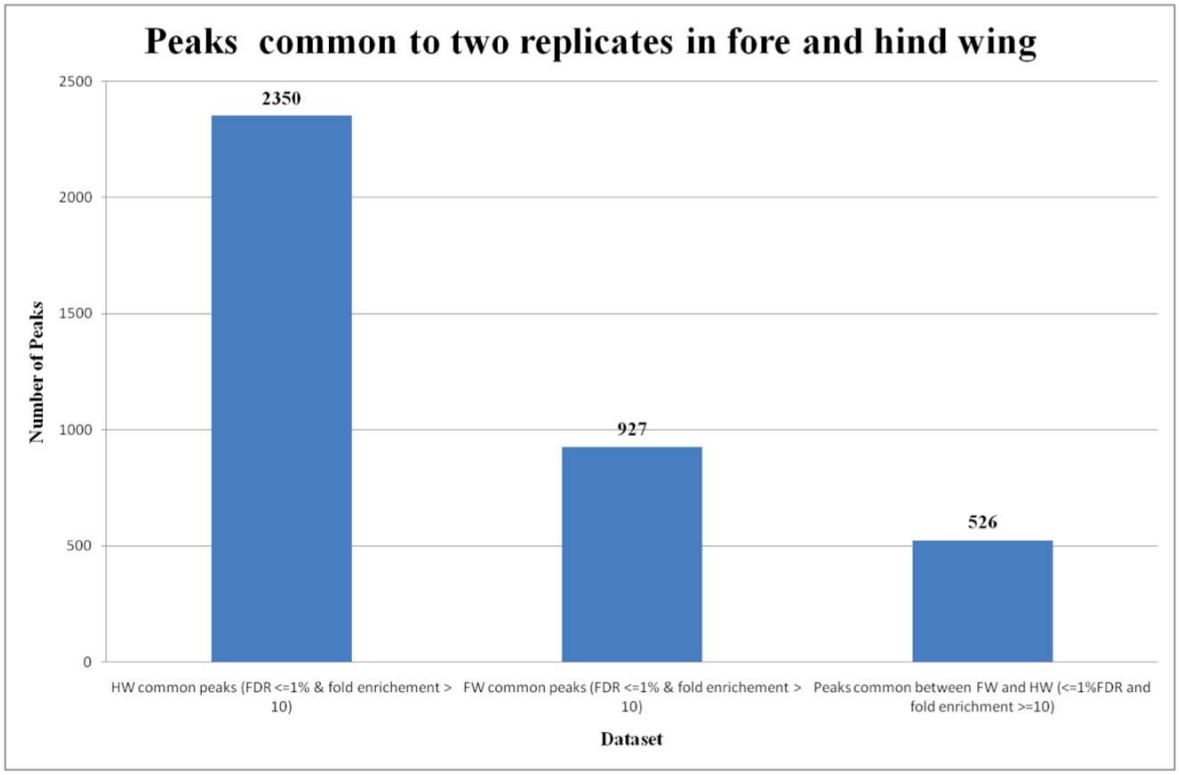


Plate 3.31: High confidence fore and hind wing peaks

Chapter 4

Comparison of direct target genes of Ubx in hind wing discs of *Apis mellifera* with those in halteres of *Drosophila melanogaster*

Introduction

Transcription factors direct gene expression by binding to promoter regions of various genes. ChIP coupled with sequencing/ microarray has been used as a technique to determine the evolutionary dynamics of transcription factor binding across different species (Borneman *et. al*., 2007; Schmidt *et. al*., 2010). Genome wide binding sites of CCAAT enhancer binding protein alpha (CEBPA) in liver of various vertebrates have been compared to test the divergence in regulatory regions amongst various mammals. Analysis of the binding sites in five different species -mouse, human, dog, opossum and chicken- revealed continuous and extensive rewiring of the gene regulatory network in vertebrates. Most of the comparison across species for regulation by a particular transcription factor revolves around comparing the motifs used by the transcription factors across various species and the genes regulated by them. A comparison of other targets of the TF, like micro RNA too is possible but it requires a detailed knowledge of the micro RNA repertoire, which is difficult to generate by in silico approach alone.

Various in silico approaches have been used for assigning a gene as a target of the transcription factor. The approach of identifying genes by various genome wide assays provides significant capabilities to study large varieties of biological mechanisms. A good gene list should contain important genes, which should be expected from the study (in the current scenario wing development genes); the list should have a reasonable number of genes; the list should have more enriched biology than any random gene list taken. There are few online softwares that check the enrichment for genes associated with particular biological process or signalling pathways in a particular gene list. Such algorithms can therefore be not only used to check the reliability of gene list but also to study the types of biological processes that have been affected.

This chapter describes how specific genes were assigned to individual peaks and the biological significance of those genes as Ubx targets. These target genes of Ubx in hind wing discs of *Apis mellifera* have also been compared to those in halteres of *Drosophila melanogaster* and a possible significance of the diptera specific Ubx target genes in shaping the haltere has been discussed.

Materials & Methods

4.1: Validation of peaks (using ChIP-qPCR)

In order to check if the peaks that were picked up in various bioinformatics analysis are true peaks and also to validate the ChIP data from first two replicates, a third independent replicate of ChIP was carried out and qPCR was performed on three peaks of both fore and hind wing. The enrichment was calculated against input and normalised using a negative control. 2X Mesa green mix from Diagenode was used for all the PCR assays while PCR primers at a final concentration of 0.2 picomoles / µl were used. The DNA template (input, chipped DNA) was used at a final concentration of 1 ng/µl. The reaction conditions and primers that were used have been described in appendix.

4.2: Assignment of genes to individual peaks

BedTools was used to search for genes near the peaks. If any gene lay within 2 kb on either side of the peak, or if the peak itself was found in a gene (in intron of the gene); the gene was considered to be a positive Ubx target. Even if multiple genes were found within 2kb of the peak on either side, all of them were considered positive hits. All the peaks that didn't have any genes within 2kb were searched for genes on either side and the first gene within 5kb on any side was taken as a positive hit (process summarised in **plate 4.3**). BioMart was subsequently used to identify the orthologs of these honeybee genes in *Drosophila melanogaster*.

4.3: Validation of few Ubx target genes in *Apis mellifera*

In order to understand the biological significance of the Ubx targets in the hind wing of honeybee, differential expression studies were carried out on the total RNA obtained from the fore and hind wing discs of honey bee larvae using real time qRTPCR. RNA isolation and cDNA preparation was done in the way already

described earlier. Briefly, cDNA from both fore and hind wing discs was diluted 1:50, 1:100, 1:200, 1:300 and 1:500 with Tris buffer, pH 8.0. Real Time reactions were set up using diagenode Mesa green Mix (2X), primers (0.2pm/µl final concentration) with the cDNA. A control of RNA (without reverse transcription reaction) was used to check the quality of cDNA and primers. Another control without template was used to check the formation of primer dimers. A house keeping gene RpL P0 was used to normalise the ct values between the fore and hind wing discs. All the reactions were carried out in triplicates. The reaction conditions for each gene and the primer sequence have been described in appendix.

4.4: Comparison of Ubx target genes in *Apis mellifera* to those in *Drosophila melanogaster*

Venny was used as an online bioinformatics tool to determine the common and specific elements between datasets. DAVID was then used as a tool to classify the genes based on their biological process and pathways. The search for biological process was limited to GOTERM_BP_FAT of DAVID. The search for pathways was by using KEGG pathways. Various datasets were fed to DAVID and a background of *Drosophila melanogaster* genes was used to calculate values like fold enrichment while other variables like percent genes associated with a particular process/ pathway were calculated based on the number of genes associated with the process/ pathway when compared to total genes.

Results & Discussion

4.5: Validation of peaks (using ChIP-qPCR)

Three peaks from the high confidence dataset of fore wing and three from high confidence dataset of hind wing were selected for validation using qPCR. There was a clear enrichment in the anti Ubx IgG chipped DNA samples as compared to the pre immune samples. Both the samples were compared against input to avoid any bias. These results, that were obtained by a different technique (ChIP-qPCR instead of ChIP-seq) on a third independent replicate of ChIP (including independent preparation of chromatin and Immunoprecipitation) show that the peaks identified by MACS are true peaks (**plate 4.1 and plate 4.2**).

4.6: Genes that were picked up as Ubx targets

A total of 1396 genes were identified as targets of Ubx in the hindwing of *Apis mellifera* of which 1182 had their known orthologs in *Drosophila melanogaster*. These 1182 genes were used subsequently for all the future analysis.

In the forewing however, only 583 genes were picked up that were the targets of Ubx, out of which 528 found their orthologs in *Drosophila melanogaster* which were used for subsequent studies (**plate 4.4).** The results have been summarised in table 3.The table also shows the common genes across various datasets. A summary of the Ubx target genes that are common to hind wing in Apis and haltere in Drosophila with their biological processes has been presented in table 4 while a summary of diptera specific target gens has been presented in table5.

4.7: Biological significance of the Ubx target genes in hind wing of *Apis mellifera*

Quantitative real time PCR on the cDNA of fore and hind wing of *Apis mellifera* showed that three out of the five genes tested were differentially expressed between

the fore and hind wing while other two were not differentially expressed (**plate 4.5**). GB15719, honey bee ortholog of fly *tousled like kinase* (*tlk*) was up regulated in the hind wing disc as compared to the fore wing disc. *tlk* in flies is known to regulate cell shape, mitotic cell cycle and chromatin organisation.

GB10329 (honey bee ortholog of fly *SmG*) was down regulated in the hind wing disc as compared to the fore wing disc of the bee. *SmG* is known to regulate mitotic spindle organisation and hence affect mitosis in fruit fly.

GB17325 (honey bee ortholog of fly *RpS30*) was down regulated in the hind wing of *Apis mellifera* disc as compared to the forewing. This down regulation was however, very subtle. The gene in flies is involved in mitotic spindle elongation and organisation.

All these genes are involved in either cell division or in regulating cell division. Another noticeable feature is that all of them show minor difference in expression levels between the fore and hind wing disc. The fact that the fore and hind wing of honeybee are only marginally different in size very well relates to the such difference in expression of these genes.

Other genes like GB13747 (honey bee ortholog of fly Dsp1) is involved in developmental process. However, the gene product was not differentially regulated between the fore and hind wing, which once again is in agreement with the identical shape of the two wings in honey bee.

4.8: Comparison of Ubx target genes associated with various biological processes and signalling pathways

Ubx target genes in fruit fly have been identified in three independent studies by Choo *et. al.*, 2011; Slattery *et. al.*, 2011 and Agrawal *et. al*.,2011. For comparing the Ubx targets in honey bee to those in fruit fly; the Ubx targets identified in the study

carried out by Choo *et. al*. and Agrawal *et. al.* were considered. A comparison of all the genes has been described in table 3. Genes that are targets of Ubx in both the hind wing of honey bee and haltere of fruit fly have been summarised in table 4; while genes that are diptera specific targets of Ubx have been summarised in table 5. The target genes identified in the study carried out by Slattery *et. al.* included a lot of non specific targets as the antibody used for ChIP studies was raised against full length Ubx protein of *Drosophila*. The other two studies were based on more specific antibodies and hence the data from these studies have been used for comparison in the following analysis.

4.8.1: Genes associated with various biological processes and signalling pathways in various datasets (Ubx target genes in halteres of diptera and hind wing of hymenoptera) (plates 4.6 & 4.7)

Genes were assigned to various Biological processes and signalling pathways by DAVID and some important processes and pathways were picked up for comparison. The comparison was carried out in two ways, in first all the genes that were targets of Ubx in hind wing disc of *Apis mellifera* or the halteres of *Drosophila melanogaster* (identified by studies carried out by Choo *et. al* . and Agrawal *et. al*.) was used and subsequently assigned to various biological processes and signalling pathways. Genes associated with various biological processes/ signalling pathways were then reported as percent or proportion of genes associated with the particular process/pathway when compared to the whole dataset in question. In plate 4.6 and 4.7, these gene proportions have been indicated as Choo *et. al.,* Agrawal *et. al*. and Ubx targets in hind wing (A.m.). The bars represent percent genes of respective dataset associated with each biological process or pathway.

Subsequently, the targets of Ubx in fore and hind wing of *Apis* were compared. A comparison of the number of Ubx target genes in fore and hind wing has been shown in plate 4.4. The comparison gave three types of genes- one that were common to fore and hind wing of *Apis*, second that were specific to fore wing of *Apis*

(not found as a target of Ubx in hind wing of *Apis*) and third that was specific to hind wing of *Apis* (not found as a target of Ubx in fore wing of *Apis*). These genes were again assigned to various biological processes and were subsequently reported as the percent genes of the respective dataset (e.g. hind wing specific targets of Ubx). In plate 4.6 and 4.7, these targets have been referred as Ubx targets specific to hind wing and Ubx targets common to fore and hind wing.

When genes associated with various biological processes in three different datasets – Ubx targets in halteres of *Drosophila* and those in hind wing of honey bee, were compared; it was found that almost all the major biological processes (only

few have been shown in the graph for sake of simplicity) were targeted by Ubx in the very ancestral insects(plate 4.6). A very similar picture emerged when the genes associated with various signalling pathways in different datasets was compared (plate 4.7).

Several processes, like wing disc development, growth and cell adhesion found a greater representation in the Ubx targets in dipterans (a greater proportion of Ubx targets in haltere of *Drosophila* were associated with these processes when compared to those in hind wing of *Apis*) (plate 4.6). These processes are essentially the ones that are related to wing patterning, suggesting that such genes are more frequently targeted in halteres as compared to hind wing disc in honeybee. The additional targets acquired in dipteran lineage by Ubx when compared to those in the hymenopteran lineage may explain the radical transformation of a wing to a haltere in diptera whereas in hymenoptera there is a slight change of forewing to a smaller hind wing in the third thoracic disc.

The proportion of genes associated with few biological processes like transcription, cell division, apoptosis were almost same for the two fly datasets and the honey bee dataset, irrespective of the difference in total number of target genes (plate 4.6). This suggests that these biological processes that are targets of Ubx in *Drosophila* were already associated with Ubx (or were target of Ubx) before the

divergence of hymenopterans and other modern insects. It is possible that the quantitative difference in the expression levels in these targets might have played a greater role in bringing about the huge change in size of haltere as compared to wing.

Almost all the major signalling pathways; except wingless signalling pathway, which has been targeted to a greater extent in halteres of fruit fly by Ubx; have been targeted by Ubx in the two insect lineages identically (plate 4.7). Almost identical proportion of Ubx target genes associated with major signalling pathways found in all the three datasets (two of diptera and one of hymenoptera) suggests that these pathways like many biological processes were targets of Ubx in the ancestral insects and a transformation of hind wing to a haltere may be brought about by the quantitative differences in targeting by Ubx of these pathways by Ubx in two lineages. Wingless signalling pathway, which has been targeted by Ubx to a greater extent in dipterans might have played a greater role in transforming a hind wing to haltere.

Overall, the picture that appears to emerge is that that *for changing hind wing to a haltere, Ubx doesn't seem to have brought any novel biological process or signalling pathway under its control specifically and dominantly in the dipteran lineage.* Though some important biological processes seem to be represented at a higher proportion in the dipteran targets of Ubx, the difference in proportion is too subtle to account for the drastic change in hind wing to a haltere. Also, few other biological processes and pathways have almost equal proportion of genes that have been targeted in dipterans and hymenopterans.

It appears therefore that, even though the number or proportions of Ubx targets that are associated with various biological processes/ signalling pathways remain the same, there may be a difference in the actual genes that have been brought under the control of Ubx in the two insect orders. In the later half of the chapter, comparison of Ubx targets in the haltere and hind wing has been described in terms of actual genes and their ontological distribution.

Another important fact that comes out from the analysis is that a large proportion of the genes that are common targets of Ubx in both the fore and hind wing of *Apis mellifera* are associated with all the key biological processes and signalling pathways (plate 4.6 and 4.7). The proportion of these genes associated with key biological processes or signalling pathways exceeds the proportion of genes associated with any biological process in any of the individual datasets- either fly alone of honey bee alone. *An important interpretation that may be drawn from this data is :these targets of Ubx are perhaps the ones where Ubx just binds to DNA – and that this binding to DNA is quite strong, but still Ubx doesn't seem to exert its effect, at least in honey bee fore wing.* There may be multiple reasons for this lack of Ubx function in the fore wing- which are discussed later.

In the honeybee fore wing bud, the peripodial membrane couldn't be distinguished from the main wing disc and the expression of Ubx appeared to be distributed throughout the wing disc. However, in the more recently evolved insects like lepidopterans and dipterans, a peripodial membrane is found and the expression of Ubx is limited to the peripodial membrane of the disc. It is therefore possible that in the course of evolution as Ubx became increasingly important for shaping the hind wing disc, expression of Ubx was restricted in a spatially different domain – called the peripodial membrane in fore wing discs.

4.8.2: Ubx target genes specific to diptera, hymenoptera & common target genes (on comparison of Ubx targets in fruit fly and honey bee) associated with various biological process and signalling pathways (Plates 4.8 - 4.11).

As the total proportion of target genes of Ubx in hind wing discs of honey bee were quite similar to those in the haltere of fruit fly, a comparison was made between the actual genes of various datasets. In the first case, targets Ubx in fruit fly, identified by Choo *et. al.* was compared to the Ubx target genes in hind wing disc of honey bee. In the second case, the targets of Ubx in fruit fly identified by Agrawal *et. al*. was compared to the Ubx target genes in honey bee. In each case, following comparison

three types of genes emerged- one that were specific targets of Ubx in diptera, second that were specific targets of Ubx in hymenoptera and third that were common targets of Ubx in diptera and hymenoptera (shown as venn diagrams in plates 4.8- 4.11). Each of these three datasets- genes specific to either hymenoptera or diptera and common to both; were associated with various signalling pathways and biological processes using DAVID and were represented as proportion of respective datasets (e.g. percent genes associated with Wg signalling pathway in diptera specific genes out of the total diptera specific genes).

Targets of Ubx in halteres of fruit fly identified by Choo *et. al.* and those in hind wing discs of honey bee are very similar in number while the targets of Ubx in fruit fly by Agrawal *et. al*., and those in hind wing disc are quite different in number; but when the datasets are compared in terms of percent genes associated with biological processes, almost identical picture emerged. In both the cases, both the diptera-specific Ubx target genes and the hymenoptera specific Ubx-target genes found a representation in all the major biological processes and signalling pathways.

Few processes like wing disc development, cell adhesion, growth and its regulation are represented at a much higher proportion in diptera specific Ubx targets as compared to hymenoptera specific Ubx targets (plate 4.8) while processes like transcription, cell cycle and its regulation, apoptosis and its regulation were represented at roughly an equal proportion in both the diptera specific and hymenoptera specific Ubx targets. Except for Wg signalling pathway, all the three other pathways that were compared, namely Notch , Hh and TGF- beta signalling pathways were targeted almost at similar proportion in both the hymenoptera and diptera specific targets (plate 4.9).

The fact that a greater proportion of diptera specific genes are associated with important biological processes like wing disc development and adhesion, and Wg signalling pathway suggests that the *novel targets that have been brought under the control of Ubx in dipterans have triggered the formation of haltere by quantitately*

increasing the differences that exist between fore and hind wing to shape a much reduced haltere. Genes involved in few processes like cell adhesion might have also modified the shape of haltere qualitatively so that it gets different shape.

A very large proportion of the common target genes of Ubx between the two species (which should also be the Ubx targets in the ancestral endopterygotes) are associated with major signalling pathways and biological processes (plate 4.8 and 4.9). It can be argued that these common or ancestral Ubx targets donot play an important role in changing the morphology of wing, as they were the Ubx targets even in the ancestral endopterygotes.

However, considering the fact that a large proportion of the genes that are common targets of Ubx in both honey bee hind wing disc and haltere disc are associated with major signalling pathways and biological processes, it can also be suggested that *the genes, once brought under the control of Ubx and if found useful have been used by the Hox gene over the long course of evolutionary history*. These genes might have been utilised similarly or in a different manner in the recently evolved insects.

As there are suggestive arguments in favour of both the ancestral and the novel Ubx targets in shaping a novel organ, it becomes necessary to estimate roughly which of the Ubx target genes- the specific target genes of Ubx (in diptera or hymenoptera) or the ancestral Ubx target genes; have played a greater role in shaping the dorsal third thoracic structure. It has been asserted in earlier chapters; that the dorsal third thoracic organ- a hind wing or a haltere is essentially a modification of the second thoracic organ or the fore wing; for which the input of Ubx is needed. If insect specific Ubx target genes in any insect order are more important in shaping the third thoracic segment than the ancestral target genes; then in a microarray study between the fore and hind wing or between a wing and haltere; those specific genes of Ubx should be represented at a higher proportion when compared to the ancestral targets of ubx and vice versa.

As no microarray study between fore and hind wing of honey bee has been reported, I decided to stick to the diptera specific genes and the common target genes of Ubx for comparison with the microarray data between wing and haltere of *Drosophila melanogaster.*

4.9: Differential expression of common &diptera specific target genes between wing and haltere (Plates 4.12 – 4.14)

In order to test relative importance of (a) genes that are conserved as targets of Ubx over the long evolutionary history, and (b) genes have come under the influence of Ubx only in dipterans, a comparison was made between microarray data and the direct targets of Ubx genes. Ubx targets obtained by Choo*et.al.* and Agrawal *et. al.* were compared to the Ubx targets in hind wing disc of *Apis mellifera*. The common targets (essentially the ancestral targets) and the Ubx targets specific to *Drosophila* (essentially the novel targets in diptera) were compared to the microarray data (joint data by Pavlopoulos *et. al.*, 2011 and Prasad *et. al*., 2006), represented in **plate** 4.12 & 4.13. The genes were reported as proportion of genes represented in microarray data out of either specific / common Ubx targets.

The graph in plate 4.14 shows that both the diptera-specific target genes and the common target genes are represented at roughly equal proportion in the microarray data. *This suggests that both the types of Ubx targets- ancestral Ubx targets and the novel target genes of Ubx in halteres have played an equally important role in shaping the haltere instead of wing.*

Huge number of diptera specific genes that do not find representation in microarray data may suggest two things- *minor differences in the expression of the genes may remain undetectable, which still may be responsible for change in morphology*. The second reason may lie also in the theory of neutral substitutions that was described earlier in introduction. These mutations as described earlier can exist in the individual even under conditions of stabilising selection and until they are detrimental to the survival of the population are not weeded out. These mutations/

DNA sequences may serve as binding sites of Ubx to DNA which for the population/ species in question is both insignificant and non detrimental. Hence these sites may continue to bind to Ubx without any consequences and hence the lower representation of these genes in microarray data.

This theory, in principle at least explains why amongst 1182 targets of Ubx in hind wing of *Apis mellifera*, 950 targets are totally specific to hymenoptera. These targets must have served as raw materials for evolutionary forces to act and only the important mutations were passed on to the newer species as motifs.

This also explains why a huge number of Ubx targets do not find any representation in the microarray data- though few of them must be essential for haltere development, but many of them must be random mutations which have been used by Ubx to bind to DNA. Some of them in the course of evolution must / may become useful targets for new species.

The fact that both the ancestral Ubx target genes and the novel Ubx target genes have played an equal role in shaping the haltere/ a novel organ in dipterans raises two questions:

1. Have the common target genes of Ubx in the two insects (or maybe the ancestral Ubx target genes) been utilised very differently in the two insects to shape different organs and if yes, how?
2. How have the diptera specific target genes helped Ubx in shaping a haltere in the fruit flies instead of a hind wing in honey bees.

4.10: Expression of important wing patterning genes (which are targets of Ubx in two insects)

To test specifically if few common target genes of Ubx have been utilised differently in the two insect lineages; expression pattern of few major wing patterning genes which are common targets of Ubx in the two insects was studied.

The expression pattern of *ct* and *sal* as shown earlier are identical in the fore and hind wing of honeybee but are different in *Drosophila melanogaster*. In the dipteran lineage, *ct* is down regulated in the posterior compartment of haltere. (summarised in plate 4.16).

There seems to be no difference in expression pattern of *sal* between the fore and hind wing bud of honeybee but there is a clear difference in the expression pattern of *sal* between wing and haltere. *Spalt* gene product is completely down regulated in the pouch of haltere in fruit flies. (summarised in plate 4.15)

Wg whose expression pattern has been studied in other hymenopteran insects (Abouheif *et. al*., 2002) show no difference in expression between fore and hind wing and a difference in expression is seen between wing and haltere. *Wg* is down regulated in the posterior compartment, in a manner quite similar to *ct.* (summarised in plate 4.17).

All these specific examples suggest that targets of Ubx in the ancestral insects have been used very differently in the recent evolutionary history to modify the wing and hence attain haltere fate. Given the fact that these genes have been continuously targeted by Ubx over 250 million years of evolutionary history, it becomes evident that these genes are very essential for wing development and have been repeatedly been used by Ubx.

As the Ubx binding itself didn't seem to make much of a difference (as shown with distinct examples where common Ubx targets in two insect orders were regulated in different ways); the mechanism of target gene selection by Ubx in hind wing discs of honey bee was studied- which is described in the next chapter.

Summary

To summarise, this chapter described how the various peaks that were identified using ChIP-seq experiments were validated. Further, the chapter also highlighted the significance of genes that are direct targets of Ubx in hind wing discs of *Apis mellifera*. On comparison of the direct target of Ubx in honeybee to those in fruit fly, it is hypothesized that neither any specific biological process nor a signalling pathway has been targeted by Ubx in dipterans to shape a haltere instead of hind wing. On comparing the ancestral Ubx targets and the diptera specific Ubx targets with the microarray data for differentially expressed genes between wing and haltere, it is proposed that Ubx has used both – its novel targets and ancestral targets to shape a haltere.

Future work in this direction needs to be done to determine how Ubx has used its pre existing targets and simultaneously employed novel targets to give rise to a new organ. As is evident from this data, a number of genes have been acquired by Ubx as its targets only in diptera- these genes must have played an important role in shifting the equilibrium towards the acquisition of haltere fate by simultaneously affecting multiple pathways and biological processes. Also, the pre existing targets of Ubx might have been utilised in novel ways to modulate various processes and pathways in different ways to shape a haltere.

While a good amount of work has been done on number of genes that are involved in wing development in *Drosophila*; a good amount of work still needs to be carried out to understand the functional relevance of various diptera specific target genes and their role in shaping the haltere. Functional characterisation of one such gene has been described in the last chapter. Also, it is essential that the role of multiple targets of Ubx in affecting various pathways be studied in the context of haltere development. Some of the diptera specific target genes might have brought the

pre existing pathways of wing development under a more stringent control leading to shaping of a very small haltere.

Plates

Chapter 4

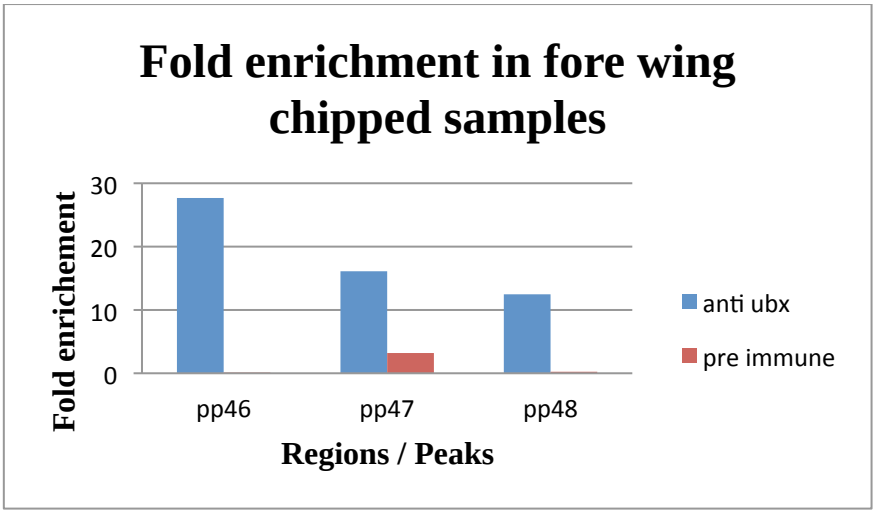


Plate 4.1: Graph showing fold enrichment for three peaks in anti Ubx and preimmune chipped DNA samples on a third replicate of ChIP.

Enrichment for both anti Ubx and pre immune samples have been calculated over input and normalised using a negative control.

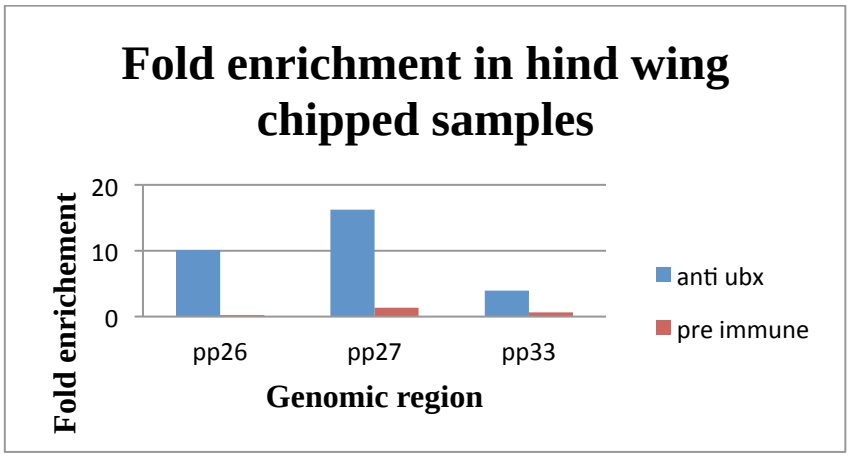
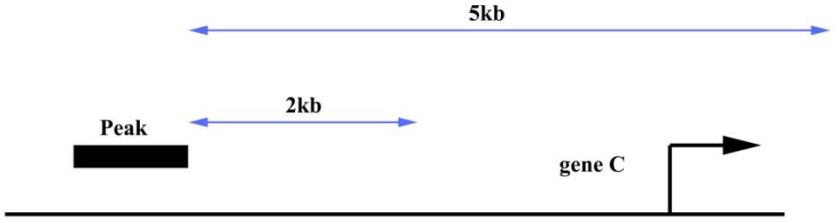


Plate 4.2: Graph showing fold enrichment for three peaks in anti Ubx and preimmune chipped DNA samples on a third replicate of ChIP.

Enrichment for both anti Ubx and pre immune samples have been calculated over input and normalised using a negative control.

genomic DNA

Both gene A & gene B were considered as targets as long as they were within 2 kb of the peaks



genomic DNA

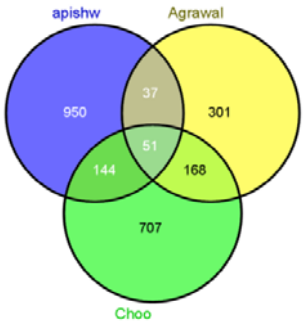
When the peak didn't have any gene within 2kb, the nearest gene on its either side within 5kb was considered as positive target

Plate 4.3: Assignment of genes to individual peaks

Total number of Ultrabithorax targets identified in *Apis mellifera* hind wing

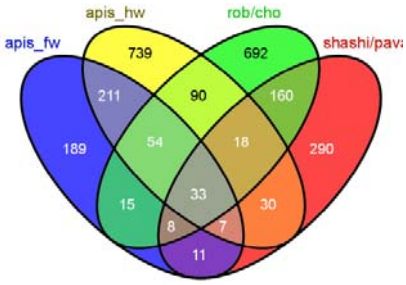
bud: 1396. Total number of Ultrabithorax targets in hind wing bud of honeybee which had their orthologs in *Drosophila melanogaster :* 1182

Total number of Ultrabithorax targets identified in *Apis mellifera* fore wing bud: 583. Total number of Ultrabithorax targets in fore wing bud of honeybee which had their orthologs in *Drosophila melanogaster*: 528



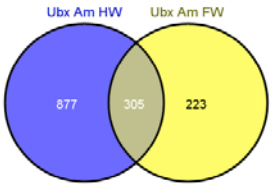
Comparison of direct targets of Ubx obtained in various studies

---



Comparison of direct targets of Ubx obtained in various studies

---



Comparison of direct targets of Ubx between fore and hind wing of *Apis mellifera*

Plate 4.4: Comparison of direct targets of Ubx

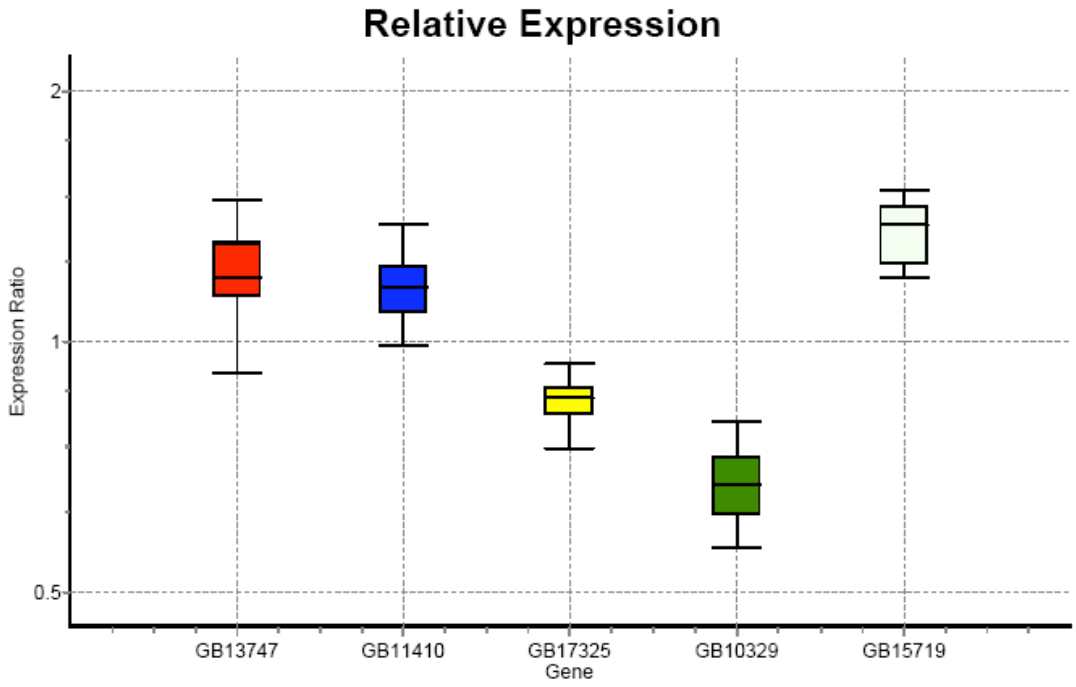


Plate 4.5: Relative expression of few Ubx target genes in hind wing when compared to forewing of *Apis mellifera*

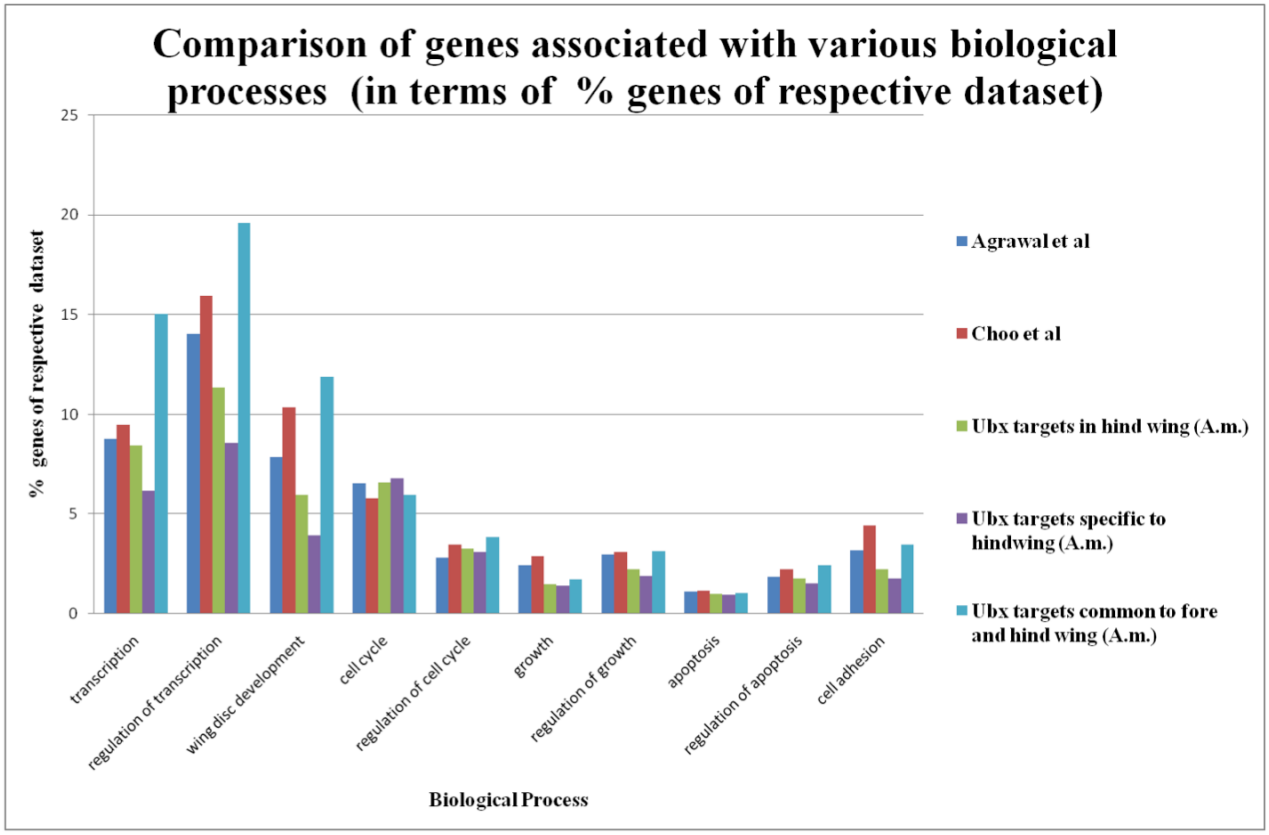


Plate 4.6: Graph comparing the genes associated with various biological processes in different datasets

Genes associated with various biological processes are reported as percent or proportion of genes associated with the particular process when compared to the whole dataset in question. In the plate, these gene proportions have been indicated as Choo *et. al.,* Agrawal *et. al.* and Ubx targets in hind wing (A.m.). The first three bars of each represent percent genes of respective dataset associated with each biological process.

Subsequently, the targets of Ubx in fore and hind wing of *Apis* were compared. A comparison of the number of Ubx target genes in fore and hind wing has already been shown in plate 4.4. The comparison gave three types of genes- one that were common to fore and hind wing of *Apis*, second that were specific to fore wing of *Apis* (not found as a target of Ubx in hind wing of *Apis*) and third that was specific to hind wing of *Apis* (not found as a target of Ubx in fore wing of *Apis*). These genes were again assigned to various biological processes and were subsequently reported as the percent genes of the respective dataset (e.g. hind wing specific targets of Ubx). In the plate, these targets have been referred as Ubx targets specific to hind wing (A.m.) and Ubx targets common to fore and hind wing (A.m.).

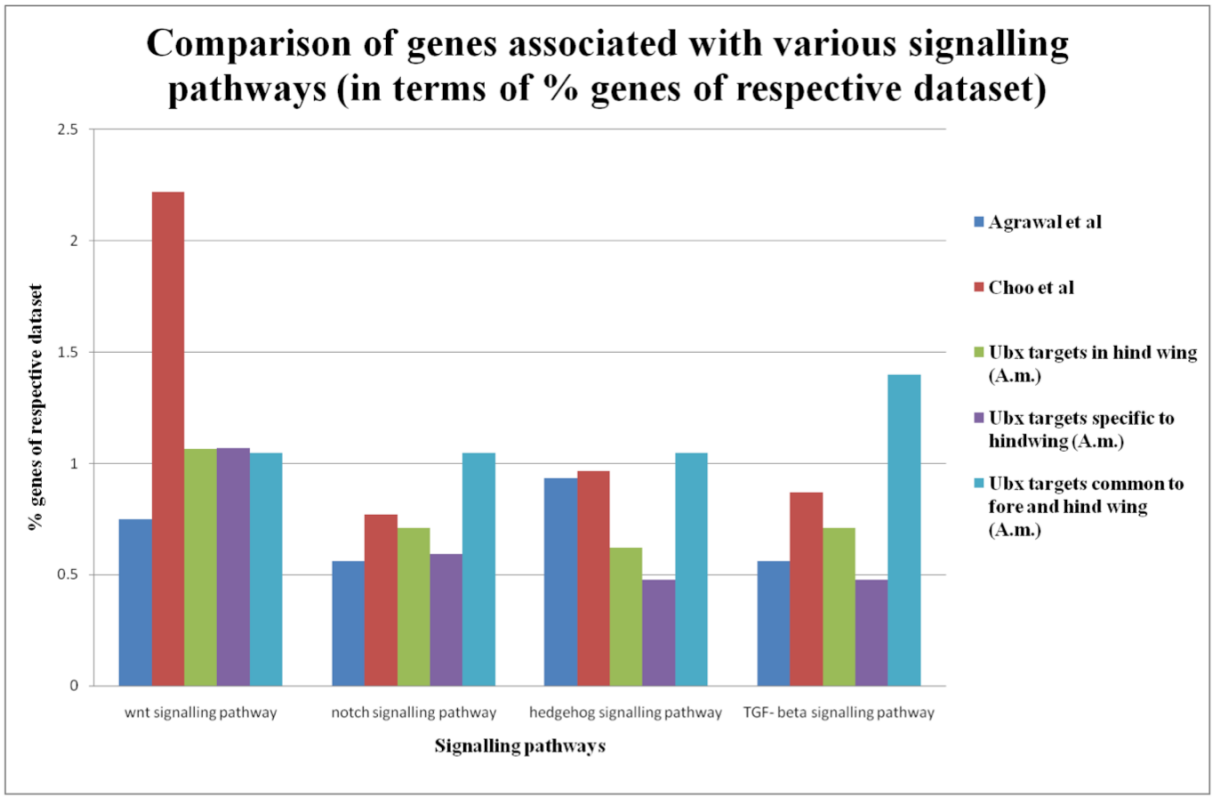
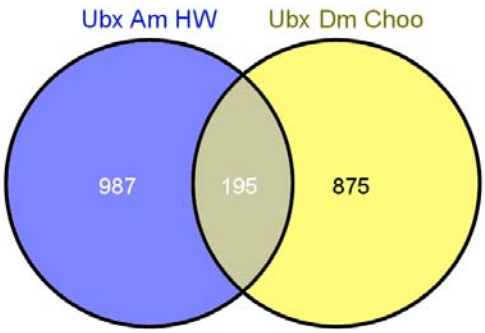


Plate 4.7: Graph comparing the genes associated with various signalling pathways in different datasets

Genes associated with various signalling pathways are reported as percent or proportion of genes associated with the particular pathway when compared to the whole dataset in question. In the plate, these gene proportions have been indicated as Choo *et. al.,* Agrawal *et. al.* and Ubx targets in hind wing (A.m.). The first three bars of each represent percent genes of respective dataset associated with each biological process.

Subsequently, the targets of Ubx in fore and hind wing of *Apis* were compared. A comparison of the number of Ubx target genes in fore and hind wing has already been shown in plate 4.4. The comparison gave three types of genes- one that were common to fore and hind wing of *Apis*, second that were specific to fore

wing of *Apis* (not found as a target of Ubx in hind wing of *Apis*) and third that was specific to hind wing of *Apis* (not found as a target of Ubx in fore wing of *Apis*). These genes were again assigned to various signalling pathways and were subsequently reported as the percent genes of the respective dataset (e.g. hind wing specific targets of Ubx). In the plate, these targets have been referred as Ubx targets specific to hind wing (A.m.) and Ubx targets common to fore and hind wing (A.m.).



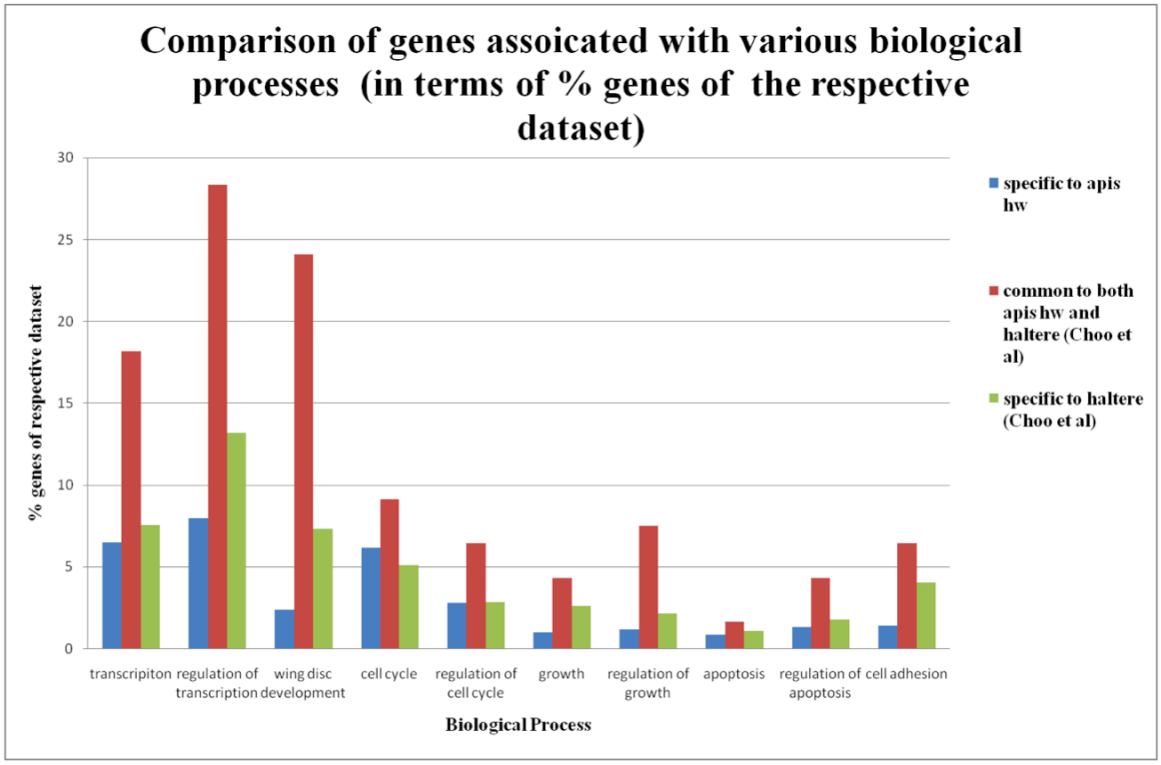
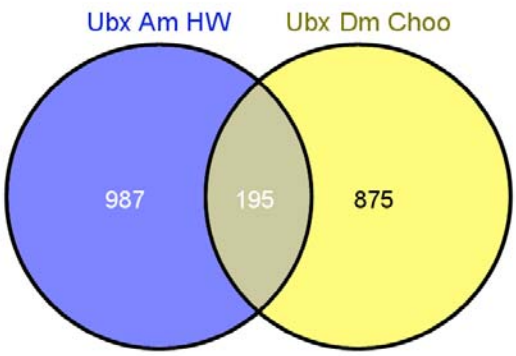


Plate 4.8: Comparison of genes associated with various biological processes. (a comparison has been made between targets of Ubx in *Apis* hind wing and targets of Ubx in *Drosophila* haltere (data taken from Choo *et. al*.,2011 and current study))

On comparison of Ubx targets in *Apis* hind wing and those in haltere (schematically shown in venn diagram above) three types of genes emerged- one that were specific targets of Ubx in diptera, second that were specific targets of Ubx in hymenoptera

and third that were common targets of Ubx in diptera and hymenoptera. For each of these three datasets, genes were taken and assigned to various biological processes and has been reported as percent/ proportion of genes associated with that biological process in terms of total genes of that dataset.



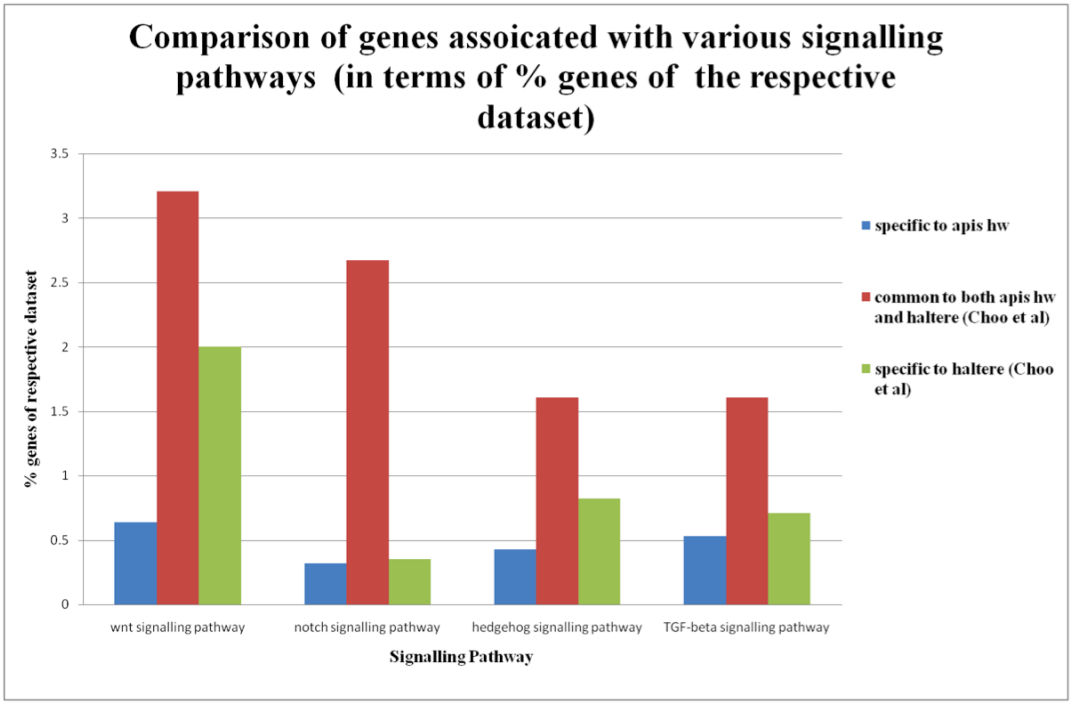
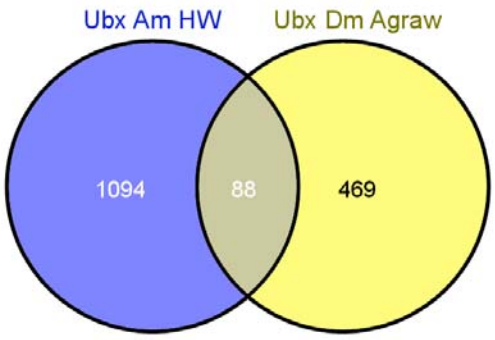


Plate 4.9: Comparison of genes associated with various signalling pathways (a comparison has been made between targets of Ubx in *Apis* hind wing and targets of Ubx in *Drosophila* haltere (data taken from Choo et. al.,2011 and current study))

Comparison and reporting of genes for the plate was done as described for plate 4.8 except that in this case genes were assigned to various signalling pathways.



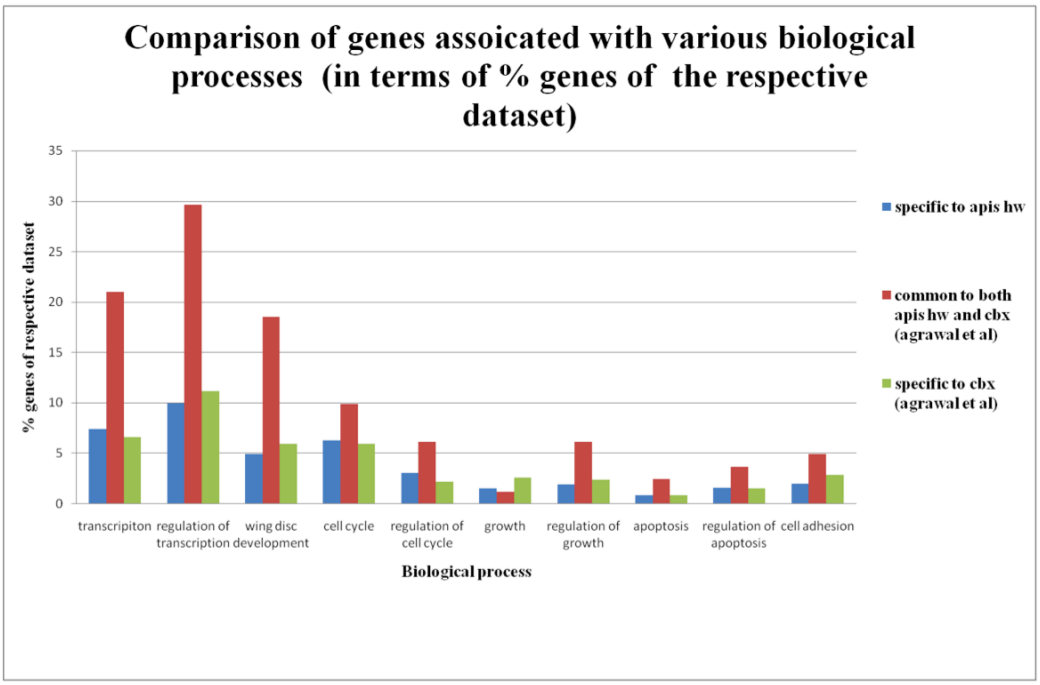
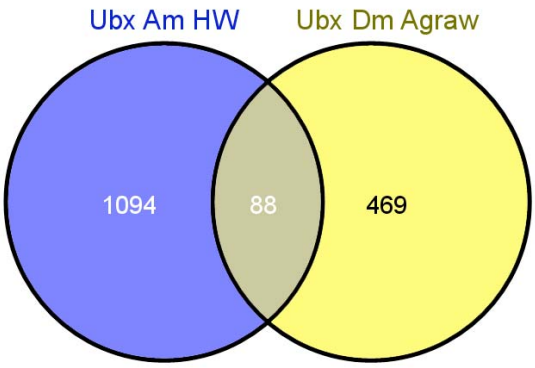


Plate 4.10: Comparison of genes associated with various biological processes (a comparison has been made between targets of Ubx in *Apis* hind wing and targets of Ubx in *Drosophila* Cbx discs (data taken from Agrawal *et. al.*, 2011 and current study)).

Comparison and reporting of genes for the plate was done as described for plate 4.8 except that in this case data from Agrawal *et. al*. was used.



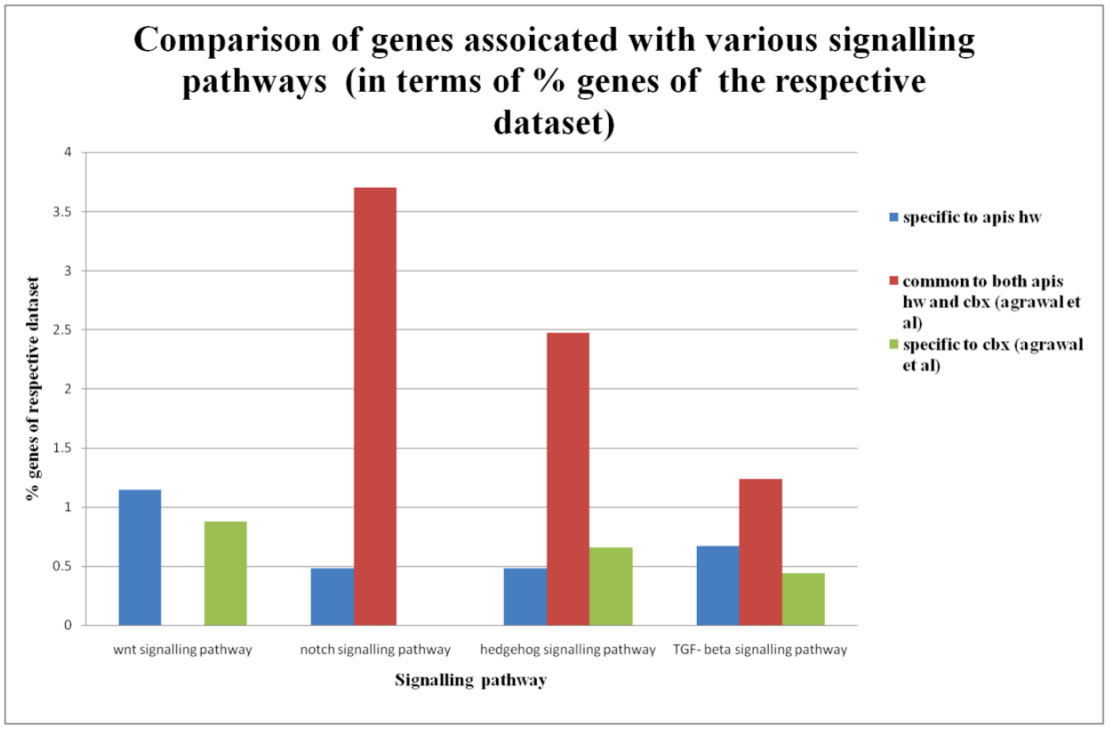


Plate 4.11: Comparison of genes associated with various signalling pathways (a comparison has been made between targets of Ubx in *Apis* hind wing and targets of Ubx in *Drosophila* Cbx discs (data taken from Agrawal et. al., 2011 and current study)).

Comparison and reporting of genes for the plate was done as described for plate 4.9 except that in this case data from Agrawal *et. al*. was used.

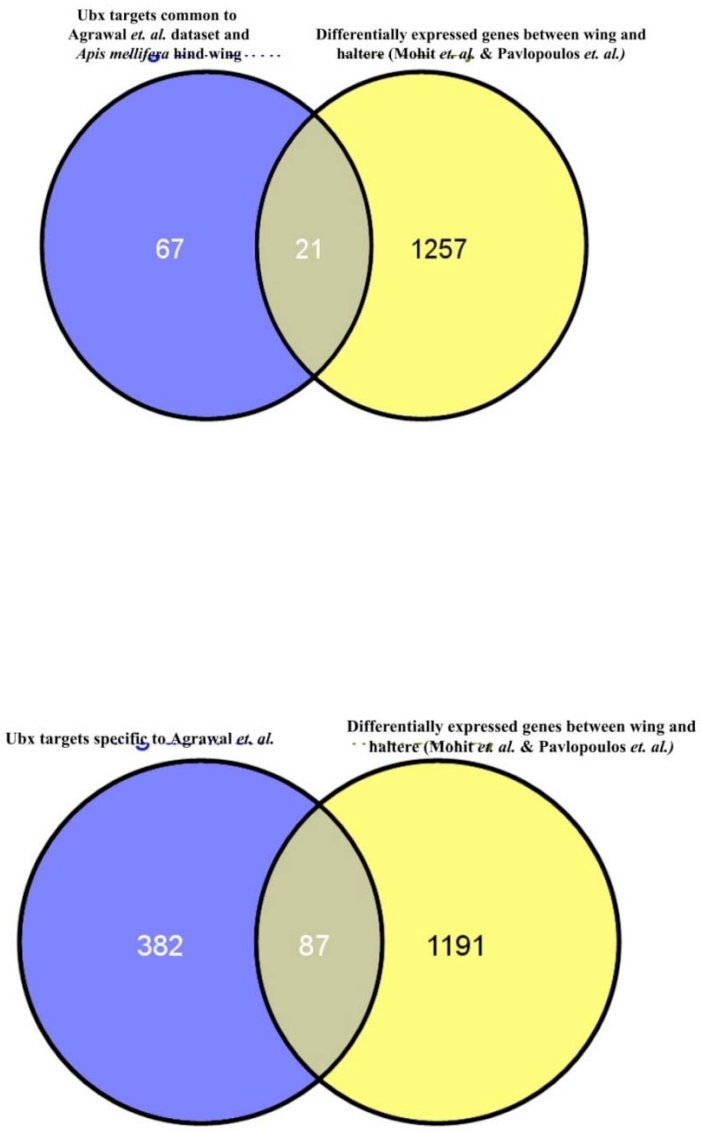
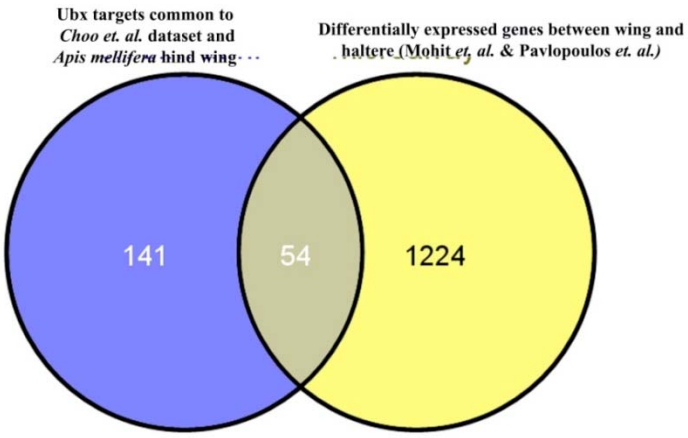


Plate 4.12: Venn diagram representing overlap of Ubx direct targets in *Drosophila melanogaster* (common to those in *Apis mellifera*/ specific to diptera when compared to those in *Apis mellifera*) with microarray data from two different studies. (Microarray data from Mohit *et. al.*, 2006 & Pavlopulos *et. al.*,2011 ; Ubx direct targets from Agrawal *et. al.*, 2011)



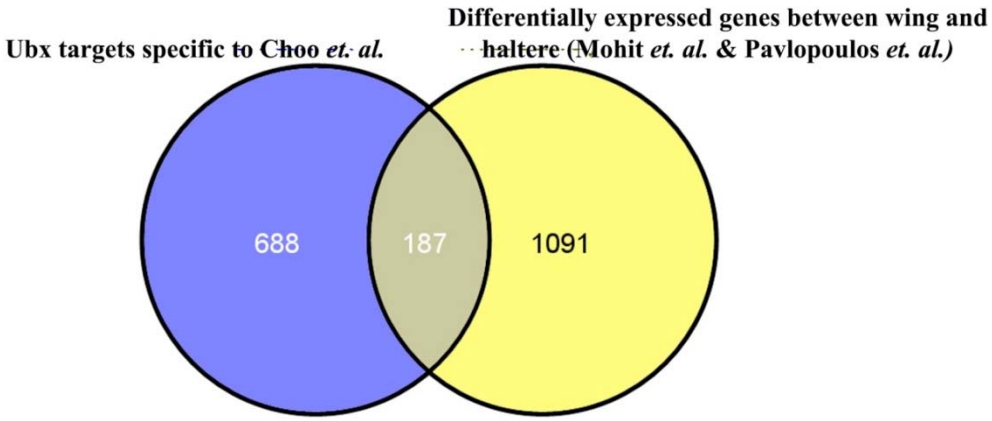


Plate 4.13: Venn diagram representing overlap of Ubx direct targets in *Drosophila melanogaster* (common to those in *Apis mellifera*/ specific to diptera when compared to those in *Apis mellifera*) with microarray data from two different studies. (Microarray data from Mohit *et. al.*, 2006 & Pavlopulos *et. al.*,2011 ; Ubx direct targets from Choo *et. al.*, 2011)

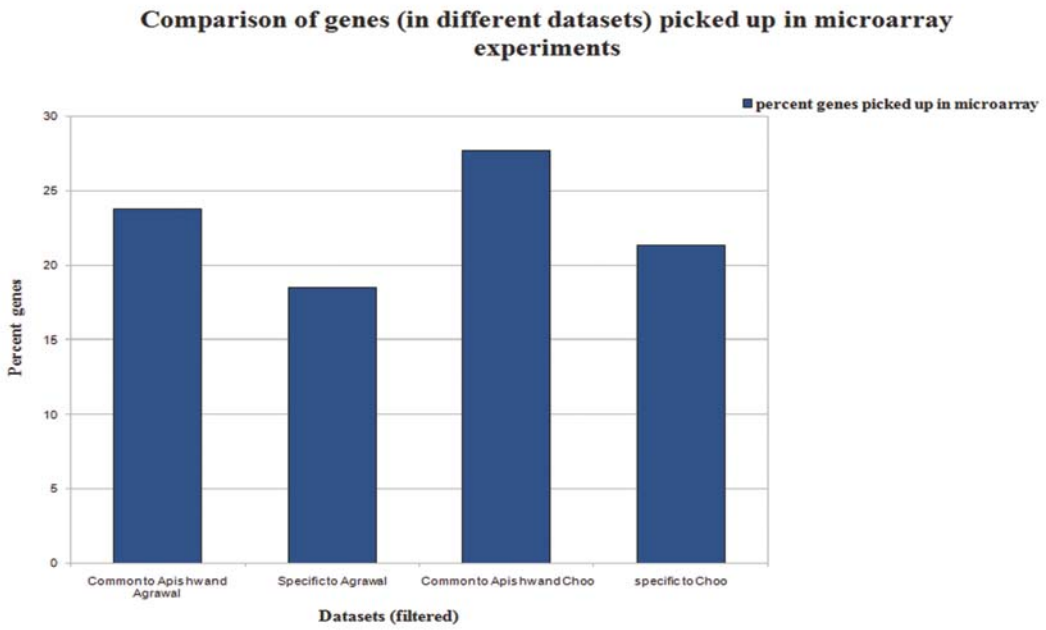


Plate 4.14: Graph representing percent genes of various datasets which have been picked up in microarray experiments

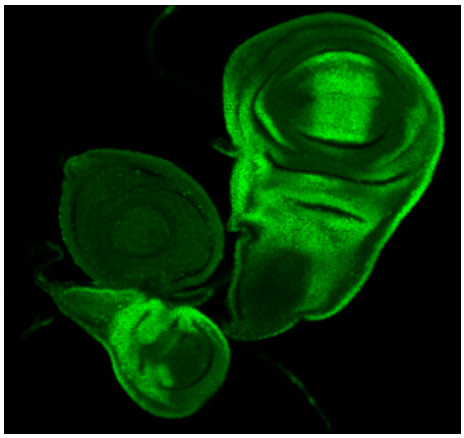


Plate 4.15 a: Differential expression of Sal between wing and haltere of *Drosophila melanogaster.* Sal is expressed along the A- P axis in the wing while it is absent in the pouch in the halteres.

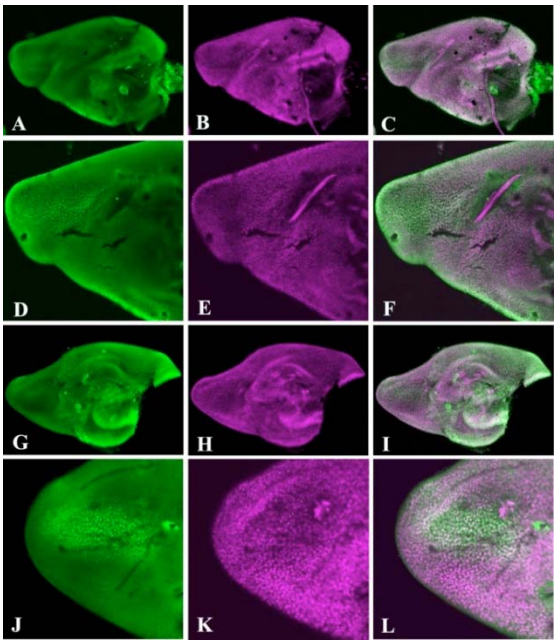


Plate 4.15b: Expression of Sal in the fore and hind wing of *Apis mellifera* (already discussed in chapter 2. The expression of Sal is identical in both fore and hind wing)

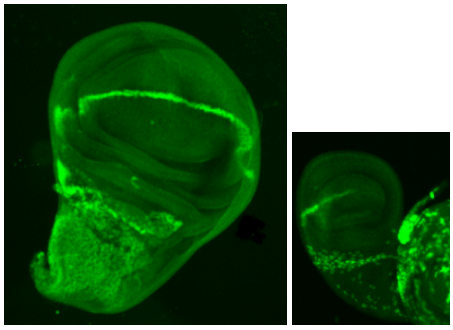


Plate 4.16 a: Differential expression of Ct between wing and haltere of *Drosophila melanogaster*. (Note that Ct is down regulated in the posterior compartment of haltere)

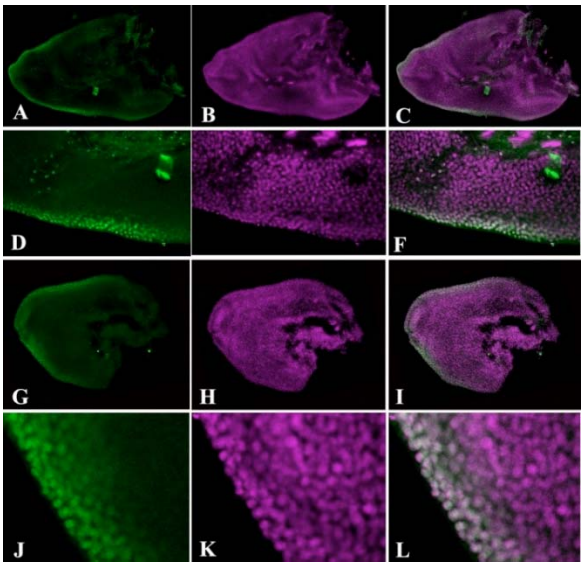


Plate 4.16 b: Expression of Ct in the fore and hind wing of *Apis mellifera*. (Note that Ct expression is similar between fore and hind wing).

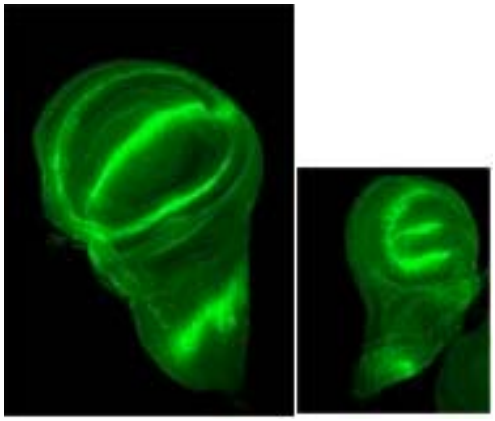


Plate 4.17a: Differential expression of Wingless between wing and haltere of *Drosophila melanogaster*. (Note the absence of wingless in the posterior compartment of haltere).

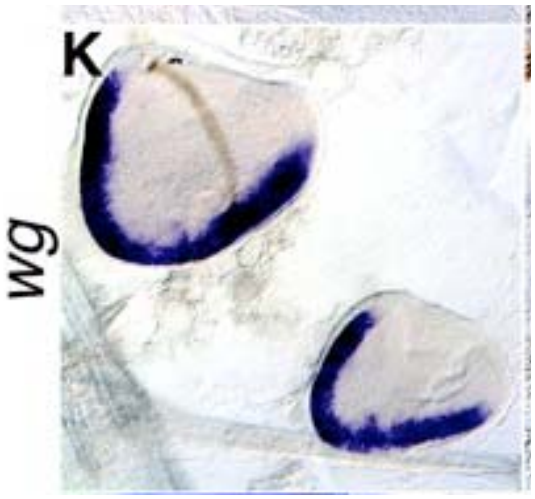


Plate 4.17b: Expression of wingless in the fore and hind wing of ants. Courtesy : Abouehif *et al.*, 2008(Please note that the expression of wingless is almost identical in the fore and hind wing discs.

Chapter 5

Comparison of motifs found in Ubx bound regions of *Apis mellifera* to those in *Drosophila melanogaster*

Introduction

Ubx is a transcription factor; it exerts its effect by binding to promoter sequences of downstream target genes. Ubx, similar to most other Hox proteins, has weak affinity to DNA binding sites and uses a core -TAAT- binding region for docking to DNA. It has been shown that the Hox proteins use a number of cofactors to bind to the DNA. As described in the previous chapter, the binding of Ubx to regulatory regions of few genes in honeybee has not resulted in any change in expression of those genes in the hind wing; but in fruit flies, binding of Ubx to the regulatory regions of the same genes has caused change in their expression pattern in the haltere disc when compared to the wing disc.

In order to understand what changes might have occurred in regulation of genes by Ubx between the two insects- honey bee and fruit fly; the direct binding sites of Ubx in honeybee were searched for motifs- or signature sequences used by transcription factors to bind to the DNA. This chapter describes the way in which the motifs were identified, and their similarity or difference with the motifs found in DNA sequences bound by Ubx in *Drosophila melanogaster*.

5.1: de-novo motif finding algorithms

Motifs are the regulatory elements, which are used by the transcription factors to bind to DNA. Identification of these short consensus sequences plays an important role in unravelling the mechanisms that regulate gene expression. One of the major problems faced by biologists and computer scientists is the discovery of patterns in DNA sequences. Given a set of sequences, one needs to find a pattern that appears frequently- however the task becomes exceedingly difficult due to higher incidents of mutations in these sequences compared to coding sequences and thus the emerging pattern might be a loosely conserved one.

Motif finding algorithms are divided into three classes based on the DNA sequence information provided.

1. Those that use promoter sequences from co-regulated genes from a single genome.
2. Those that use orthologous promoter sequences from a single gene of multiple species (phylogenetic footprinting).
3. Those that use promoter sequences from co-regulated gens as well as phylogenetic footprinting.

Based on the algorithm that the motif finding program uses, they may be classified into two major types:

1. Word based (string based) methods
2. Probabilistic sequence model based methods

The former method is based on counting and comparing oligonucleotide frequencies. The method guarantees global optimality and they are appropriate for short motifs and therefore are useful for eukaryotic genomes. Some of the commonly used algorithms are YMF, MITRA, Weeder and WINNOWER.

In the Probabilistic method, the model parameters are estimated using maximum likelihood principle or Bayesian inference. Probabilistic methods are more appropriate for motif finding in prokaryotes, where the motifs are generally longer than eukaryotes. The methods require fewer search parameters but rely on probabilistic methods of regulatory regions. Nested MICA and MEME are some of the algorithms that use the probabilistic method.

Tompa *et. al.* assessed performance of thirteen motif finding programs, namely AlignACE, ANN-Spec, Consensus, GLAM, Improbizer, MEME, MITRA, MotifSampler, Oligo/dyad analysis, QuickScore, SesiMCMC, Weeder and YMF. They generated datasets of DNA sequences containing known binding sites. When different programs were compared it was observed that Weeder outperformed other

motif finding algorithms. MEME and YMF too did better on few datasets. The study indicated that complementary tools for motif finding should be used for identification of motifs and the motifs found by *de- novo* methods should be validated by other means. Keeping this in mind- two motif finding programs – MEME and Weeder (which were the best programs found in the study and based on two different algorithms- probabilistic and word based respectively) were used in the study. The motifs obtained were then pooled and the representative motifs were reported and validated.

TRANSFAC database was used to identify motifs that are previously reported as binding sites for various transcription factors. Program like MATCH was used from TRANSFAC to find out the motif occurrences in DNA sequences. A detailed description of each of these programs has been provided in the following section.

Materials & Methods

5.2: Identification of motifs from chipped DNA sequences

Motifs in the chipped DNA sequences of *Apis mellifera* were identified using MEME and Weeder. MEME algorithm was developed by Bailey and Elkan (Bailey *et. al*., 1994) for identifying motifs in unaligned biopolymer sequences. MEME is based on three ideas for discovering motifs- subsequences that actually occur in the biopolymer sequences are used as starting points for EM algorithm to increase the probability of finding globally optimum motifs; it removes the assumption that each sequence contains only one occurrence of the shared motif; and finally a method for probabilistically erasing shared motifs is incorporated so that several distinct motifs may be found in the same set of sequences.

Weeder (Pavesi *et. al.*, 2004) uses a word-based approach to identify motifs in a given set of sequences. In this case different oligonucleotides recognised by a transcription factor are described by their consensus, representing for each position the nucleotide that appears most frequently in the binding sites. All the oligonucleotides that do not differ from their consensus in no more than a given number of nucleotides can be assumed to be bound by the same transcription factor.

The chipped DNA sequences were masked using Repeat masker and motifs were searched in masked sequences too. Motif identification was done on two data sets- the first dataset contained all the peaks/ genomic sequences (detected at high confidence as described in earlier chapters) bound by Ubx while the second dataset was generated by selecting 200 random peaks out of all the peaks detected. Both datasets were subjected to motif analysis using MEME and Weeder (both masked and unmasked sequences were analysed separately with slight difference in one parameter namely motif length).

Only significant motifs (motifs with e value <= 0.01 in MEME and the most significant motifs reported by Weeder) were chosen for subsequent analysis. STAMP was then used to align the motifs and representative motifs of each family was reported and used for further analysis.

5.3: Calculation of motif frequency in the chipped DNA and random DNA fragments

BedTools and Samtools were used to generate 300 files of random DNA fragments of *Apis* genome, each containing 2400 fragments of length 1000bp. MATCH was used to determine the frequency at which the motifs occurred in the random as well as in chipped DNA sequences. The option of minimising false positives was used to get data with higher confidence. Fold enrichment of motifs was calculated by dividing the occurrence of motifs in the chipped DNA fragment over that in random DNA sequences.

Results & Discussion

5.4: Mechanism of gene regulation by Ubx in *Apis mellifera*

Ubx being a transcription factor, regulates downstream genes by binding to specific DNA sequences. These short DNA sequences are called motifs as described earlier. Using two different algorithms to identify motifs, eight motifs with high confidence were identified. These motifs have been summarised in **plate 5.1**. Known motifs identical to these motifs were also searched for and the best matching motifs found in TRANSFAC database. The results have been summarised in the plate. Two motifs didn't have any similar motifs in the database.

Ubx binding motifs that have been reported by earlier studies and the motifs obtained by ChIP studies on *Drosophila* followed by similar searches have been summarised in **plate 5.2 and 5.3**.

Only one of the identified motifs namely motif1weeder5 seemed to contain A/T rich sequences, which could be used by homeodomain containing proteins for binding. The motif also found good similarity to PEND protein- a homeodomain containing protein in plants. Inability to identify AT rich known Ubx binding motifs using de novo motif search methods, may be due to the background genome which itself is highly AT rich (60% of honeybee genome). This also reflects weakness in all motif-discovery algorithms.

Other motifs showed good similarity to binding sequences for GAGA, MAZ, E2F, Adf-1, c-Myc proteins.

5.5: Enrichment of the motifs in the chipped sequences as compared to the random genomic fragments

Ubx motifs, from all the earlier studies (Ekker *et. al.*, Noyes *et. al.* and Mann *et. al.*) were under represented in the chipped DNA sequences as compared to the random

DNA sequences. The data is in agreement with earlier study carried out in *Drosophila* (Agrawal *et. al.*, 2011). The distribution of Ubx motifs in both fruit fly and honey bee shows that the motif has been found quite uniformly across the genome right from very ancestral insects. However, a number of other motifs like GAGA, Adf-1 are enriched in the chipped sequence by up to 5 fold. Other motifs like MAZ and E2F were enriched by about 3 fold in the chipped DNA sequence of honeybee when compared to the background DNA sequence. The enrichment of various motifs in the chipped sequences in *Apis mellifera* has been summarised in **plates5.4, 5.5 and 5.6**.

As far as TAAT core sequence is concerned, similar distribution of motifs was seen in *Drosophila*. There was no difference between DNA sequences pulled down by Ubx antibodies and sequences randomly pulled out from the database. *This is consistent with the previous observations that Ubx may loosely associate itself on the DNA throughout the genome. These associations may need a number of co factors, so that Ubx may stably bind to the location and exert its effect.* It is by this mechanism by which Ubx may attain specificity in regulating its targets. The use of cofactors also allows the protein to have variability in its effects- both spatially and temporally.

5.6: Preference in utilisation of motifs

The hymenopteran lineage diverged from the other insects around 250 million years ago and subsequently gave rise to multiple species one of which is *Apis mellifera*. As discussed earlier, neutral mutations seem to arise in all organisms and remain in the organism until they start becoming detrimental. The so called neutral mutations may be even selected based on the advantages they confer to the population.

Over the course of evolutionary history, a large number of mutations must have arisen in the lineage leading to *Apis* and *Drosophila* and the ones that we find today; some of which are used by various transcription factors to bind to the DNA sequences; may have arisen as random mutations during evolution. However, the fact

that many of them are used by both *Drosophila* and *Apis*, two species whose lineage has diverged quite some time back- suggests that these mutations/ motifs are very ancestral ones and have been used by Ubx since then. This is also in agreement with the functional conservation of Ubx in the insect orders. However, as there has been a change in suite of target genes between *Apis* and *Drosophila* (some of them being conserved while others have changed); there is a good possibility that Ubx has used some of these motifs preferentially in one lineage over the other while regulating these genes.

As discussed in the previous chapter, as many as 238 genes are common targets of Ubx in honey bee and fruit fly. A number of mutations are likely to have accumulated over 250 million years in both the lineages independently. This would lead to multiple situations. One, certain motifs are so important for Ubx function, they would occur at similar frequency amongst genes that are targets of Ubx in either *Apis* or *Drosophila* or both. Certain new motifs may occur in either insect lineages as they evolve with new targets of Ubx or new mechanism of regulation by Ubx.

In order to investigate, which motifs have been utilised preferentially in honeybee as compared to fruit fly and vice versa; the frequencies of motifs in various datasets were tested. Frequency of motif occurrence in the chipped sequences of honeybee and Drosophila was calculated. For each motif, this frequency was calculated for the sequences that corresponded to genes common to both honey bee and *Drosophila* (these would be the genes that were ancestral targets of Ubx) as well as for the sequences that corresponded to genes specific to either *Drosophila* or *Apis*(this should be the genes that are novel targets of Ubx in either insect lineage). Motif frequency in the sequences (corresponding to common genes) was divided by motif frequency in the sequence (corresponding to specific genes) to get a ratio, which would give an idea about its utility in shared genes over specific genes (result summarised in **table 6**).

A ratio of 1 is suggestive of the motif being a very important for Ubx in binding or activity- as this motif is shared equally in the genes that are either ancestral targets of Ubx or specific/novel targets of Ubx in the insect lineage in question. A ratio of less than 1 for a motif suggests negative selection for this motif when it is associated with common/ancestral Ubx targets in the insect lineage in question. A ratio greater than 1 suggests positive selection for this motif when associated with common/ancestral Ubx targets in the specific insect lineage.

We observed that except for few motifs *Apis* and *Drosophila* sequences showed different patterns of representation of motifs, which is expected considering 250 million years of divergence. Interestingly, the Ubx-binding core sequence motif, which is not enriched in the chipped sequences when compared to random sequences in both *Drosophila* and *Apis*, appear to be over represented in the ancestral targets of Ubx in *Drosophila*, suggesting its employment for Ubx function in the ancestral Ubx targets in the insect. Under-representation of this motif in *Apis* (in the ancestral targets when compared to the novel targets) suggests that many ancestral targets of Ubx may have lost Ubx control, even when Ubx continue to bind to their regulatory regions. It is possible that binding may have become weaker in those targets.

A very similar distribution of motifs was noticed for TCF and it can be suggested that TCF too has been employed as a cofactor of Ubx in regulating the ancestral Ubx targets differently in the dipteran lineage.

Few motifs like GAGA, Snail, delf/ grainyhead and MAD are found to be uniformly distributed in both the types of sequences (chipped sequences corresponding to both the ancestral and common targets of Ubx) in honey bee. This is suggestive of the fact that these motifs or the proteins that bind to these sequences are very important for the binding or action of Ubx. Hence for any chipped DNA fragments, may it be *Apis* specific or common between *Apis* and *Drosophila*; these motifs are found with equal frequency- essentially they have to be present to impart stability to Ubx- DNA interactions.

If these factors are important for binding / stabilising the binding of Ubx to DNA, they should be present at equal frequencies in other insects too - and this is what exactly what was found for *Drosophila melanogaster*. All these motifs were uniformly distributed in chipped sequences corresponding to both the ancestral and specific targets of Ubx in *Drosophila melanogaster*.

Few motifs like MAZ, Myc, Adf and E2F have a ratio greater than 1 for honey bee sequences. This suggests that these motifs became very important in hymenopteran lineage for exerting the effect of Ubx in the ancestral Ubx genes. It is also possible that the presence of these motifs or the proteins binding to these motifs, dilute the repressive effects of Ubx and hence only a slightly smaller hind wing is seen in honey bees. Just the reverse seem to happen in case of dipterans, where these motifs have been lost from ancestral Ubx targets which suggests that Ubx might have used different factors in these sequences to exert its effect in *Drosophila.*

Overall, it seems that while few motifs like GAGA, MAD, Snail and dElF-1 have remained very important for Ubx action throughout the insect lineage, various other motifs have been used in different ways to achieve various outcomes.

5.7: Mechanism by which Ubx bound genes may have been utilised differentially in the two insect orders

Previous microarray studies suggest that while majority of the targets of Ubx are down-regulated in the Drosophila haltere, few are upregulated too. However, when the microarray targets were examined for the enrichment of motifs that are associated with repressors or activators, no clear consensus emerged. In order to understand the specific effect of Ubx on its target genes, the direct binding sites of few important Ubx regulated genes were identified and compared.

5.7.1: Comparison of Ubx binding sites in the promoter region of *sal* gene in two species (Plate 5.7)

Sal is expressed in the wing pouch of *Drosophila melanogaster* along the AP axis in response to Dpp gradient *(Nellen et.al.,* 1996). The expression of *sal* is silenced in the haltere pouch by the action of *Ubx (*Weatherbee *et. al.,* 1998*)*. In the wing discs of *Apis mellifera* and other hymenopterans studied till date, *sal* is expressed in the wing disc (again along the AP boundary). Its expression however, is identical in both the fore and hind wing disc of honeybee (plate 4.15, current study) and in other hymenopteras (Abouheif *et. al*., 2002). *Sal* is expressed also in the notum and along the hinge of the wing disc in both dipterans and hymenopterans. However, this expression is identical between the wing and haltere and hence not being studied.

Sal has been identified as a direct target of Ubx in *Drosophila melanogaster* in a number of earlier studies (Makhijani *et. al.,* 2007; Galant *et. al*.,2002). The current study identified *sal* as a direct target of Ubx even in honeybee. As its expression is identical in the hind wing and fore wing disc, it is possible that binding of Ubx to this target in the ancestral insects had only subtle effect on Sal expression. It is possible that dipteran lineage has acquired binding sites for other cofactors making the binding of Ubx more effective.

In order to explore this possibility, regions bound by Ubx in *Drosophila* and *Apis* corresponding to *sal* enhancer were searched for motifs/ binding sites for known transcription factors. Two direct binding sites for Ubx were pulled down from honey bee genome that are upstream of *sal* gene- one is around 30 kb upstream, while the second is about 2.5 kb upstream. In *Drosophila*, the wing blade specific enhancer of *sal* to which Ubx also binds lies around 9 kb up stream the promoter. As the complete gene of *sal* in *Apis* is not yet annotated (only CDS is known in honey bee), the DNA that corresponds to 30 kb upstream was chosen for comparison to the Ubx binding region of *sal* enhancer in *Drosophila.*

On comparing the two sequences, it is observed that some motifs are present in both honey bee and fruit fly, while few are found specifically in either of them. Some motifs such as Ubx, TCF, Scalloped, Caudal and Brcz1 are present in both, while motifs such as Kruppel, Ttk, Snail, Dorsal, Paired and Zeste are present specifically in *Drosophila*. Others such as GAGA, MAZ, MAD, Adf1, CF2 and Knirps are present in *Apis* only.

5.7.2: Comparison of Ubx binding sites in *ct* locus of *Apis* and *Drosophila* (plate 5.8)

Another gene that is a known target of Ubx in halteres of *Drosophila* is *ct*. The gene is also a direct target of Ubx in hind wing discs of honey bee but as shown earlier, there is no difference in expression pattern between the fore and hind wing disc. The direct binding sites of Ubx in *ct* locus for the two insect species- were compared to identify the specific factors/ motifs that have arisen in fruit fly, which may have led to the repression of *ct* in haltere discs.

The direct binding site of Ubx in *ct* locus of *Apis* was identified using bioinformatics and subsequently only one site was chosen which showed considerable similarity to DNA sequence of *ct* locus of *Bombus terrestis* (another hymenopteran species). The 1868 bp conserved region of DNA which was bound by Ubx was located 23.9 kb downstream of the *ct* start site. Motifs were identified in the conserved region using MATCH program of TRANSFAC. Condition to minimise false positives were given so that only very significant motifs were picked up.

Similarly, direct binding site of Ubx in *ct* locus of *Drosophila melanogaster* was identified using bioinformatics (Choo *et. al*). The direct binding site of Ubx lay approximately 12 kb downstream of the transcription start site. The difference in the distance from the start site is reflective of the fact that the total span of *cut* gene in *Drosophila* is around 70 kb and that in *Apis* is around 375 kb. Two conserved sequences (500 bp apart) were identified in the region that are bound by Ubx in *Drosophila* (conservation essentially means nucleotide conservation across various

dipterans). MATCH was used to identify the motifs in these two regions. Condition to minimise false positives were given so that only very significant motifs were picked up.

Once again multiple motifs were picked up in both species, namely- Dorsal, Caudal, HSF, DeAF, SGF, MAD, TCF and ABDb. Many motifs are present only in honeybee sequences viz. CF2, ADF1, GAGA, ANTP, Paired, Apterous, Myc, and E2F. Some of the motifs that were picked up only in *Drosophila* sequences are binding sites for Snail, BRCZ2, ABDa, Eve, Engrailed.

The most interesting ones out of these is Engrailed. The protein has been shown to act as a co factor of Ubx in compartment specific repression of target gene *dll* in embryos. As the product of gene *ct* is repressed only in the posterior compartment, it is likely that Engrailed acts as a cofactor of Ubx in suppressing this gene.

5.7.3: Comparison of Ubx binding site in *wg* locus of the two insects (Plate 5.9)

Wingless (wg) is another target of Ubx in the halteres and is suppressed only in the posterior half of the haltere pouch. The expression of this gene product remains identical between the fore and hind wing of honeybee - shown in this study and study on other hymenopterans (Abouheif *et. al*., 2002)

To have a closer look at the changes that might have taken place in the Ubx binding region that made the repression by Ubx possible in dipterans, the binding sites of Ubx in the two insects were compared.

Binding site of Ubx in the promoter region of *wg* gene in *Apis* revealed by ChIP-seq experiment shows very high homology with a region upstream of wingless in *Bombus terrestis* (another hymenopteran) and hence was considered for this

analysis. MATCH was used to identify the motifs in the region using a criteria to minimise false positives.

The *wg*-enhancer region bound by Ubx was identified based on the data reported by Choo *et. al.* and motifs were identified in the region by MATCH.

Once again a number of motifs such as Ubx, Dorsal, Apterous and E2F are common to both *Apis* and *Drosophila*. However, each of the insects showed a good number of motifs that were specific to them. The motifs for the binding of Hairy, Zeste, Dref, GAGA Deaf, Mad, Paired, Apterous, knirps, Myc and MAZ were found specifically in *Apis* sequence, while the motifs such as Ftz, Snail, Brcz2, Elf1 SGF3, Zen, Engrailed, Tcf , Brk, Caudal, ABDa and ABDb were found only in *Drosophila* sequence.

The presence of Engrailed binding motif once again only in Drosophila reflects compartment specific repression of *wingless*.

Snail and BRCZ2 are two more transcription factors besides Engrailed whose motifs are present in both *wg* and *ct* locus of *Drosophila* but are absent in honeybees. As the expression pattern of these two genes in the wing pouch are similar, it is possible that Snail and Brcz2 too function with Engrailed and Ubx.

As described above, motifs for the binding of GAF, Paired, Apterous and Myc are found only in *Apis* sequences of both the genes, but are absent in *Drosophila* sequences. These may provide an additional or alternative mechanism for the observed repression of *wg* and *cut* in haltere discs, i.e. absence of these motifs are important for Ubx-mediated repression.

Summary

On comparing the motifs found in Ubx-bound regions in hind wing discs of *Apis* to those found in the Ubx bound regions in halteres of *Drosophila*, it is observed that Ubx uses similar motifs for target gene selection in both the insect orders. Yet few of the target genes common to *Apis* and *Drosophila* are not identically regulated. Closer examination of the Ubx-binding sites of such genes suggested that Ubx modulates the activity of its target genes by associating with various co-factors in a context specific manner, which are brought in the vicinity of Ubx during the evolution of various cis-regulatory elements.

Future work in this direction needs to be done to determine which motifs have become instrumental in regulating the pre existing targets of Ubx in a different way in dipteran lineage. Various reporter constructs of the promoter regions bound by Ubx in hymenopteran lineage can be made and their expression in fruit fly should be studied. Specific motifs should then be introduced or removed so that the honey bee sequence bound by Ubx becomes quite identical to its *Drosophila* ortholog; the expression of which in *Drosophila* should be then analysed to see which motifs have actually become more important in the dipteran lineage to shape a haltere.

Plates

Chapter 5

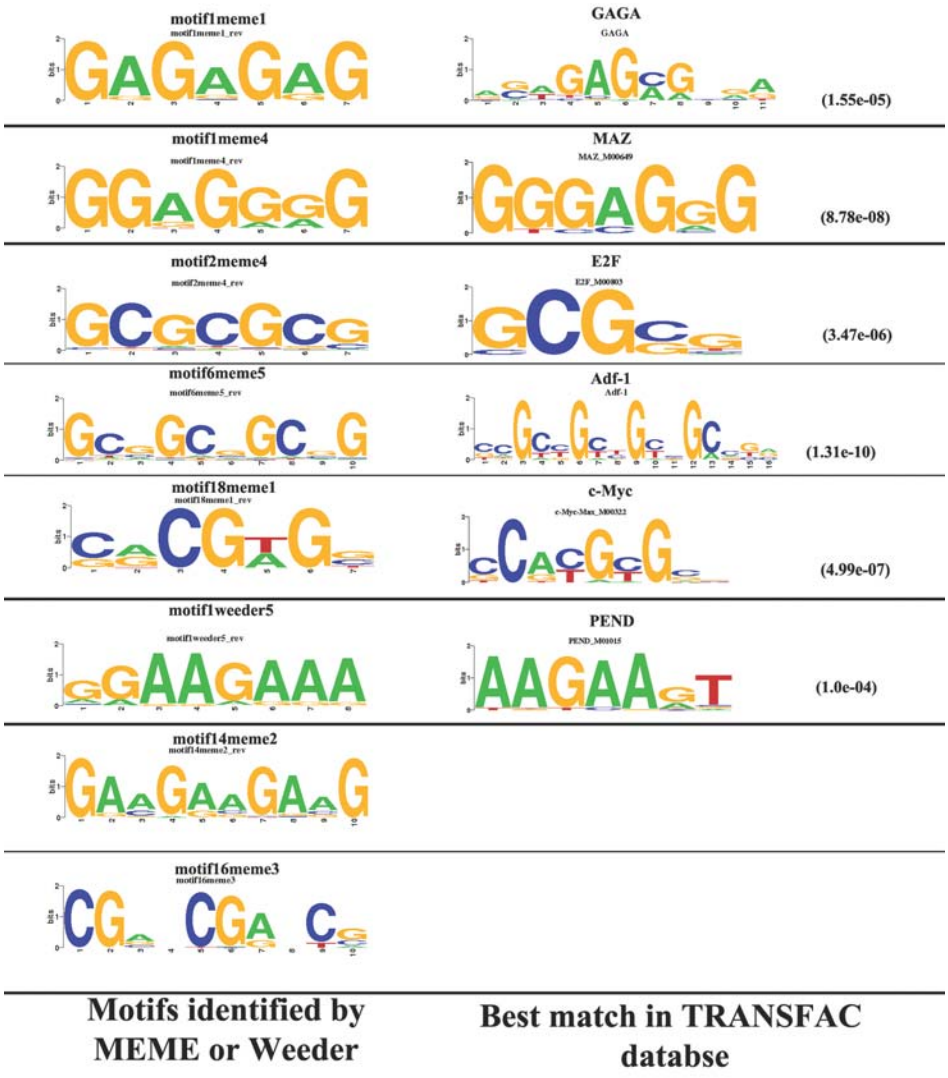


Plate 5.1: Motifs identified by different de novo motif discovery tools

The similarity to known motifs listed in TRANSFAC database has been shown.

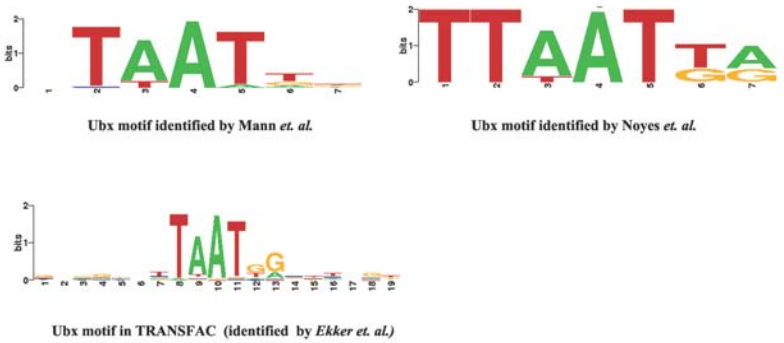
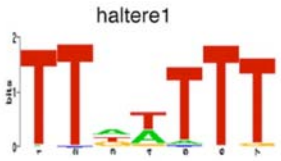
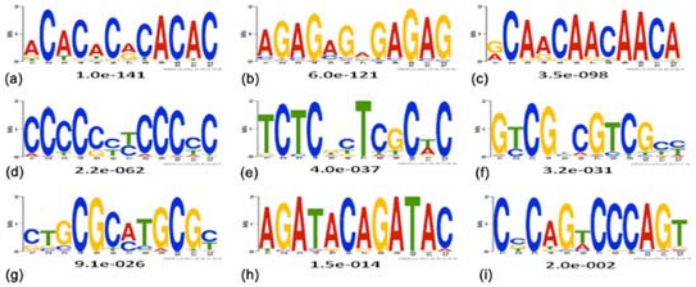


Plate 5.2: Well known Ubx binding motifs (identified by different studies)



Ubx binding motif reported by Choo *et. al*, 2011



Motifs discovered by *de- novo* motif search on sequences pulled down by Ubx protein (Agrawal *et. al.*, 2011)

Plate 5.3: Motifs enriched in Ubx chipped DNA in the haltere and Cbx discs (as reported by respective studies)

Motifs identified by MEME or Weeder	Fold enrichment of motif in chipped sequences over random sequences
	6.117
	3.587
	4.857
	5.075
	1.988
	1.631
	4.03
	2.584

Plate 5.4: Fold enrichment of motifs identified in this study

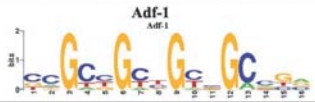
TRANSFAC motifs that are similar to the de novo identified motifs	Fold enrichment of motif occurence in Chipped DNA seq when compared to random sequences
GAGA 	6.711
MAZ 	3.373
E2F 	2.763
Adf-1 	6.048
c-Myc 	1.993
PEND 	0.658

Plate 5.5: Fold enrichment of TRANFAC motifs in the chipped sequences of honey bee

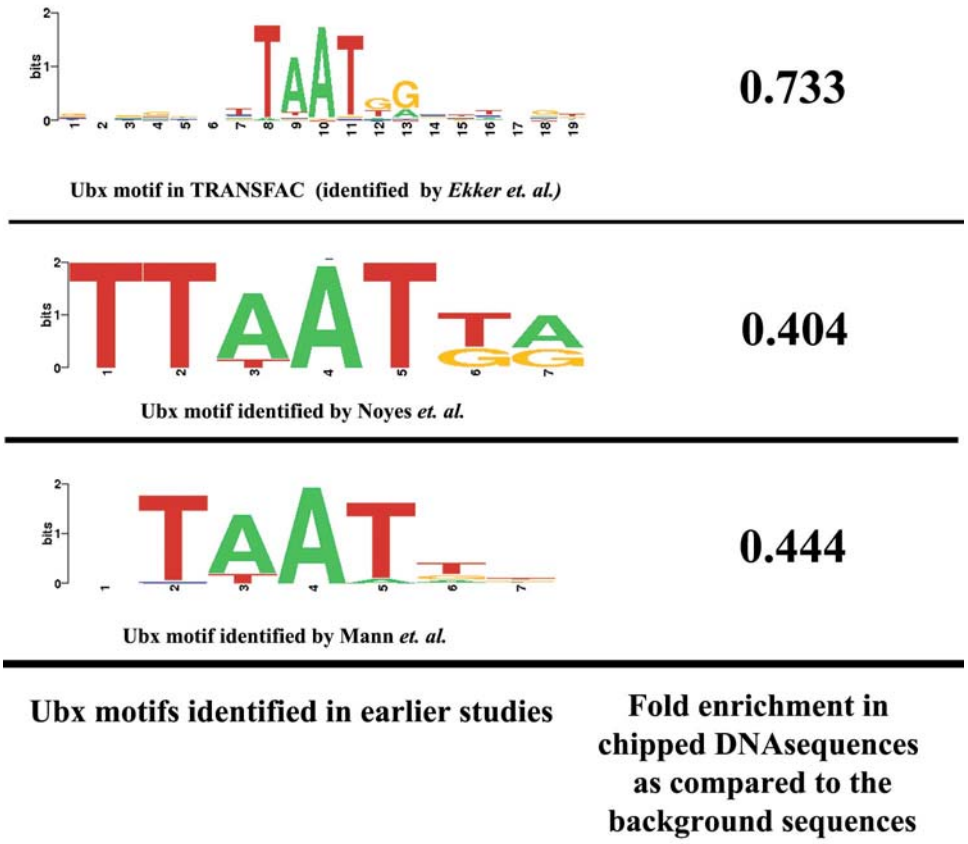
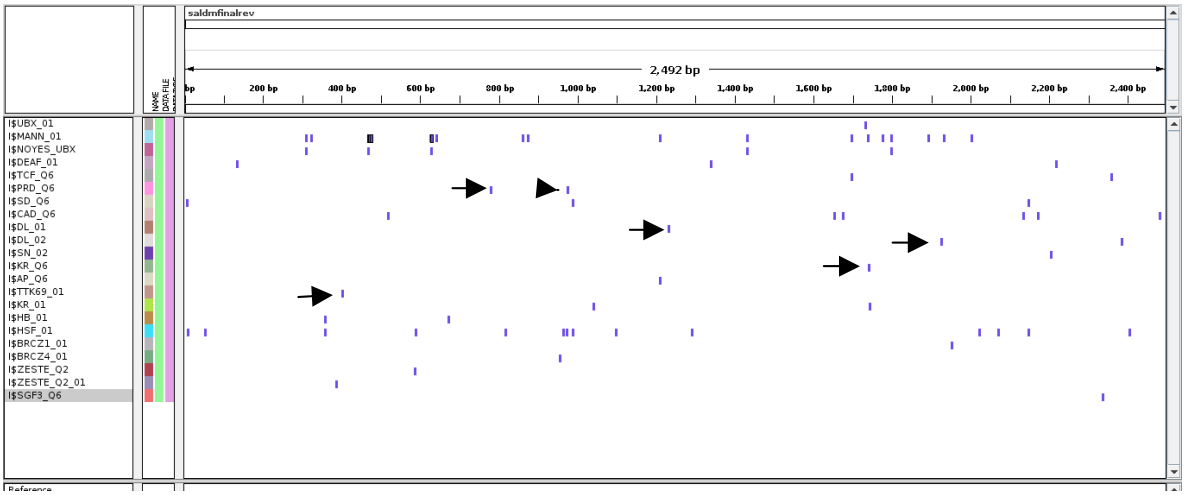
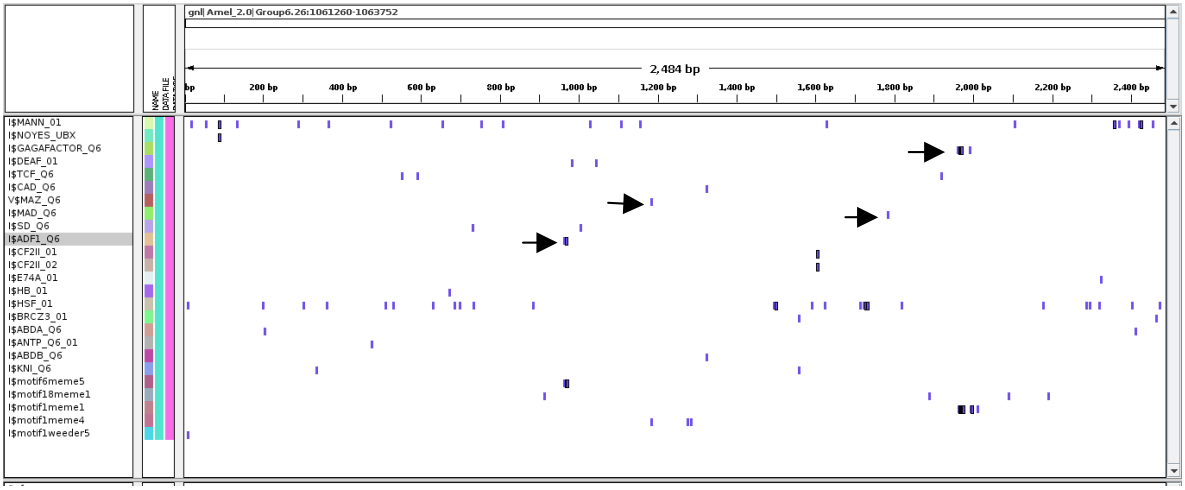


Plate 5.6: Fold enrichment of Ubx motifs (identified in various studies) in the chipped sequences of honey bee

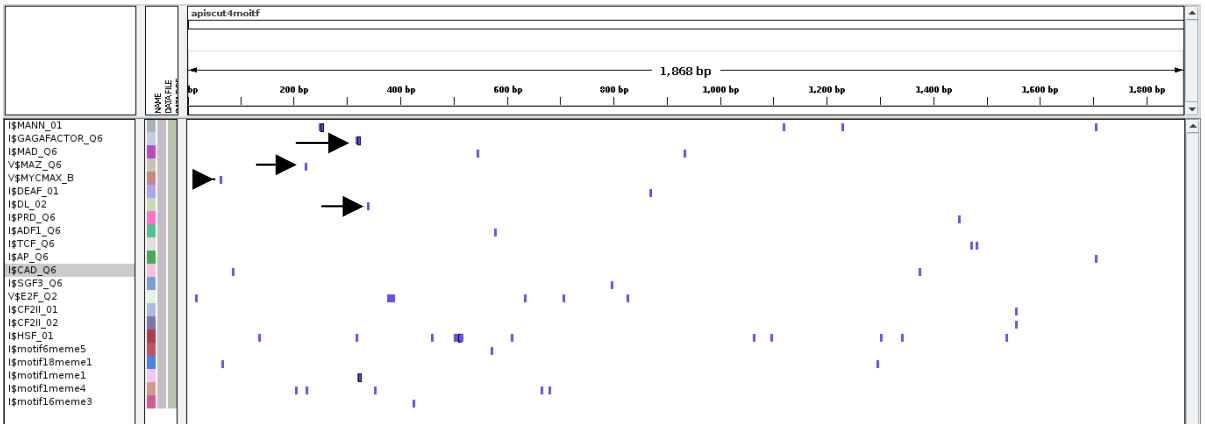


Spalt enhancer element in *Drosophila melanogaster* bound by Ultrabithorax

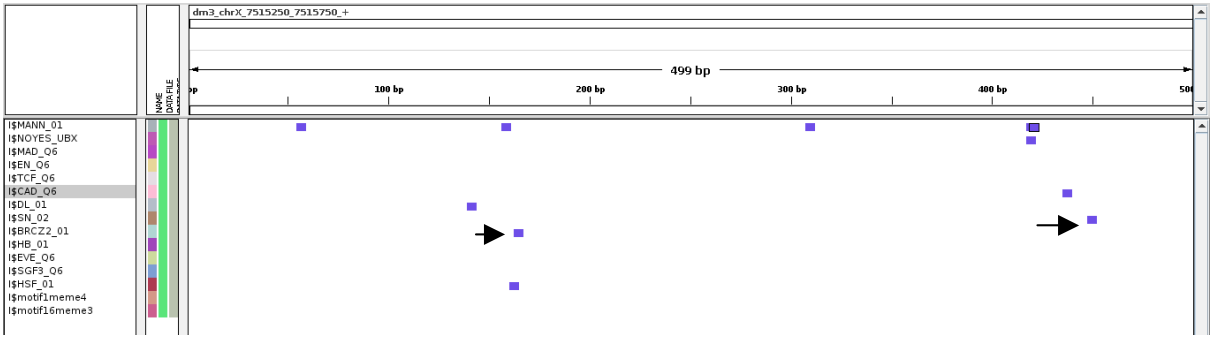


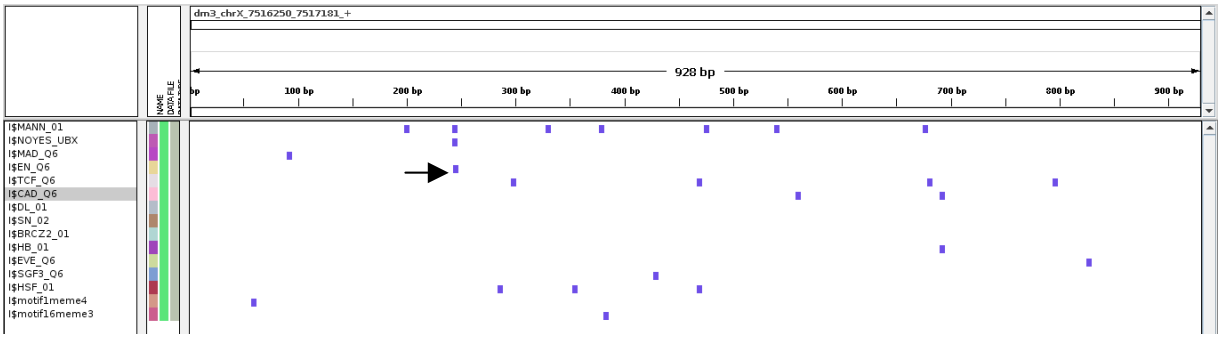
Enhancer element of *Spalt* bound by Ultrabithorax in *Apis mellifera*

Plate 5.7: Comparison of direct binding sites of Ubx in *Spalt* locus of *Drosophila melanogaster* and *Apis mellifera*



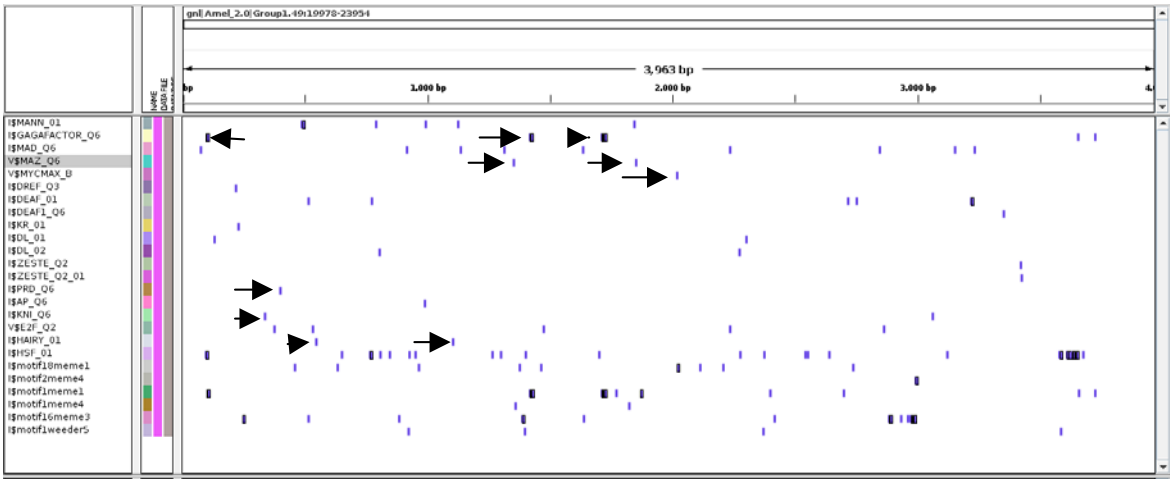
Motifs identified in DNA sequence bound by Ubx (intronic region of *cut* gene in *Apis mellifera*)



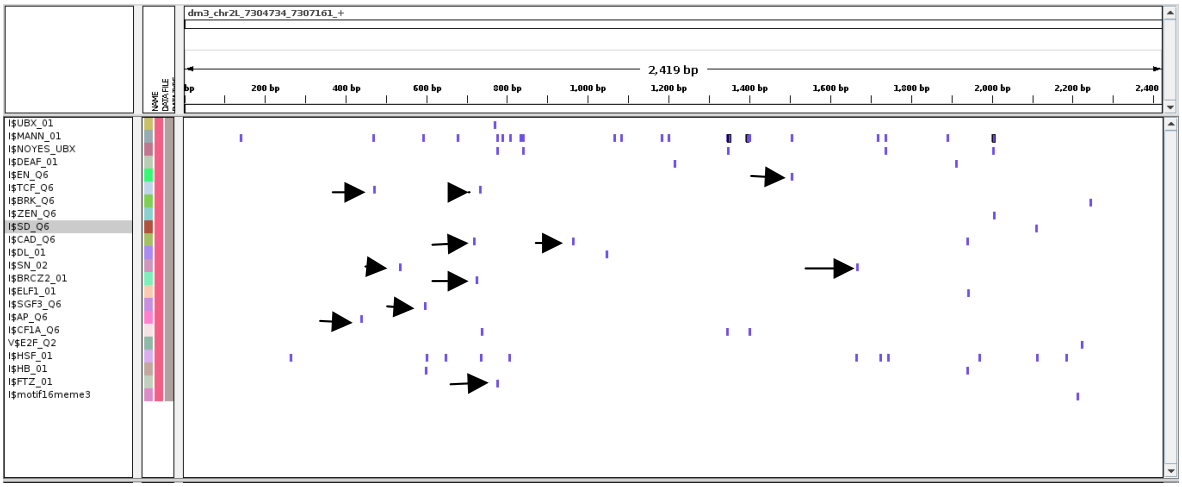


Motifs identified in DNA sequences bound by Ubx (intronic region of *cut* gene in *Drosophila melanogaster*)

Plate 5.8: Comparison of direct binding sites of Ubx in *Cut* locus of *Drosophila melanogaster* and *Apis mellifera*



Motifs identified in DNA sequence bound by Ubx (promoter region of *wingless* gene in *Apis mellifera*)



Motifs identified in DNA sequence bound by Ubx (promoter region of *wingless* gene in *Drosophila melanogaster*)

Plate 5.9: Comparison of direct binding sites of Ubx in *wingless* locus of *Drosophila melanogaster* and *Apis mellifera*

Chapter 6

Characterization of *Gliolectin*, a diptera-specific target of Ubx

Introduction

As discussed in the previous chapters a large number of targets of Ubx are common between *Drosophila* and *Apis*, although not necessarily have the same effect on their expression patterns in the two insect species. Ubx also targets few novel genes to shape the haltere in dipterans.

This chapter describes the characterisation of one such gene, *Gliolectin (Glec),* which is a diptera specific target of Ubx and is also differentially expressed between the wing and haltere. The role of down regulation of this gene in shaping the haltere is also discussed.

6.1: Modification of a wing to haltere by Ultrabithorax

Earlier studies have shown that haltere is a modification of the wing and is shaped by Ubx by modulating the wing development plan (Makhijani *et. al.*, 2007; Pallavi *et. al*., 2006; Prasad *et. al.*, 2006; Bajpai *et. al*., 2004; Prasad *et. al*., 2003). It is therefore important that the role of targets of Ubx target in wing development must be understood to have a better insight of the effect of modulation of their expression in halteres. In the earlier chapter, targets of Ubx in two different insect orders have been compared. A list of all the genes that are targeted by Ubx specifically in diptera and the genes that are targets of Ubx in both hind wing of *Apis mellifera* and in haltere of *Drosophila* have been provided in the previous chapter.

While several genes were studied for their expression patterns, *Glec* was selected for detailed functional characterization. The expression pattern of Glec has been described earlier (Prasad *et. al.*, 2006). It is expressed along the D-V boundary in wing imaginal discs and absent in haltere discs. Bioinformatics analysis and earlier studies show that it is a single pass transmembrane protein that binds to carbohydrate moieties As many of the genes organising the wing shape that are present at the D-V boundary are glycoproteins, and these genes (such as *wg*) are down regulated in haltere; any protein that is capable of binding to these proteins and expressed in a

similar pattern would be a good candidate for characterization. *Glec* that can bind to carbohydrate residues and is expressed along the D-V boundary was therefore chosen for further analysis.

6.2: Wingless signalling pathway

Wg functions as a morphogen in shaping the identity of wing along the D-V axis. It is a secreted glycoprotein that elicits cellular responses by binding to frizzled-family of membrane receptor complexes. These complexes elicit three types of responses: (1) Beta catenin (Armadillo) mediated Wg signalling (Logan *et. al*., 2004); (2)Planar cell polarity (Seifert *et. al*., 2004); and (3) Calcium related signalling which targets adhesion and other processes. In the absence of Wg signalling, cytoplasmic Arm is recruited to a complex that assembles around a scaffolding protein called Axin, where it is phosphorylated at its N terminus by Glycogen synthase3 beta (GSK3 beta) .The phosphorlylated Arm is targeted for degradation via proteasome, which keeps the cytoplasmic concentration of Arm at low levels. Upon signalling by Wg, a fraction of the cytoplasmic Arm is stabilised, probably by modification and enters the nucleus. Interaction of Arm with TCF family of proteins helps in transcription of downstream target genes. Dishevelled plays an important role in modulating the activity of Axin based destruction complex so that the interaction of Wg with its receptors leads to stabilisation of Arm. Wg signalling therefore acts as inducer to set up multidimensional transcription states. It also acts to stabilise the transcriptional state (Reviewed by Hayward *et. al*., 2008) (**Plate 6.1**).

6.3: Interaction between Wingless and Notch signalling pathways

Wg pathway also interacts with a number of other pathways like BMP, Hh, Ras/RTK pathways. Its interaction with Notch signalling pathway, however appear to be prevalent under various developmental contexts. Notch signalling was first identified in the context of lateral inhibition during the context of peripheral nervous system (PNS) development in insects. In *Drosophila*, the initial events that set up PNS development depend on Wg signalling and are followed by lateral inhibition by

Notch. Essentially, Wg mediates pre patterning and Notch mediates the inhibitory process (Martinez Arias, 2002).

Various events leading to pattern formation in the wing disc has been discussed in Chapter 1. Along the D/V axis, mutations in *wg* and *Notch* in the wing disc mutually enhance each other's effect, suggesting that Wg and Notch signalling pathways can act synergistically with each other. The explanation to this enhancement lies in the Gene Regulatory networks (GRNs) that drive wing development.

6.4: Notch signalling in defining vein/intervein boundaries

Notch signalling is often used in enhancing differences and generating a switch like behaviour during cell fate determination of neighbouring cells. One such process where Notch refines cell fates is the vein/intervein boundaries in the wing blade. The wing vein boundaries in *Drosophila* are sharply defined by a process involving Notch signalling. *Delta* or *Notch* loss of function genotypes result in vein thickening of varying intensities. *Delta* locus itself was first identified on the basis of thickenings of the distal termini of longitudinal wing veins. *Notch* gain of function however leads to vein shortening (**plates 6.2 and 6.3** show various phenotypes obtained by two independent groups when either *Delta* or *Notch* was down regulated in the pupal stage or mitotic clones of *Delta* mutants were analysed in the wing).

6.5: Vein development and morphogenesis in *Drosophila*

In *Drosophila*, there are four longitudinal veins formed by proximo-distal stripes of cells that appear more compact and have higher pigmentation than the intervein cells. The position of these veins is determined during the third instar larva by the expression of *Veinlet (ve)*(Sturtuvent*et. al*., 1993*).Ve* encodes a membrane protein with seven transmemebrane domains (Bier *et. al.*, 1990). The protein facilitates signalling via *Torpedo*, *Drosophila* homologue of EGFR. High levels of EGFR signalling occurs in places of Ve expression and this activity directs the wing cells to

adopt a vein fate. *Ve* is required to activate *Delta* transcription and *Notch* represses the transcription of *Ve*, therefore inhibiting vein development wherever it expresses. Transcription of Notch is higher in broad domains that correspond to interveins, while Delta is restricted to veins. The asymmetry in distribution of Notch and Delta results in activation of Notch at the vein/ intervein boundary and separation of veins from inter veins. In boundary cells, within the domain of *Delta* expression, transcription of *Enhancer of split (Espl)* is increased, which maintains *Notch* expression. Localised expression of Ve is a critical component of vein development. Activation of Notch signalling using *Ax* mutations leads to loss of Ve signalling while loss of Notch signalling leads to increase in number of cells expressing *Ve*. Essentially increase in Notch signalling during larval instar leads to repression of *Ve* expression and decrease in the activity of Notch leads to increase in Ve activity. In discs with reduced levels of Ve, there is no expression of *Delta* but the levels of Notch remains unchanged (de Celis *et. al.*. 1997).

The veins are determined during larval stage of development while they develop and differentiate during the pupal stage. Reduction in the level of either *Delta* or *Notch* in this stage would lead to decrease in neurogenic signalling which in turn lead to over commitment of vein cell state and hence lead to the formation of thickened veins. It is important to note that the clones of *Delta* and *Notch* form vein tissue only when they are adjacent to normal vein tissue.

During the second phase of cell layer apposition, cells in the intervein region begin to extend basal processes and only the vein cells remain unapposed. Signalling at this stage blocks the response of intervein cells to the cues that specify the vein fate and hence only these cells appose and form the intervein. In the third larval instar, *Delta* is expressed almost in the entire wing pouch though maximum expression is seen in stripes along the prospective wing margin and future longitudinal veins. *Notch* too is found in the entire wing pouch but its expression is lower in cells where Delta is present. 6 hours APF, after the eversion of wing disc, when the dorsal and ventral surface of wing pouch appose, the expression of *Delta* still remains quite like

in the third instar larval stage. Notch expression during this stage is more refined and appears more in the intervein cells adjacent to the delta expressing cells. At around 30 hours APF, the vein cells have a very high concentration of Delta and Notch is expressed highly in the flanking cells (Genetic interactions have been summarised in **plate 6.4**).

As described below, Glec appears to be a component of both Wg and Notch pathways.

Materials & Methods

6.6: Fly maintenance

Fly stock were maintained on standard corn flour and agar media supplemented with yeast and malt. Canton S (cs) flies were used as wild type flies in all the different studies and were maintained at 25ºC. The composition of fly food has been described in appendix.

6.7: Genetic crosses

Various mutations were balanced and different combinations of mutations were brought together by using standard genetic techniques. Standard protocols for fly crossing were carried out- virgin females were mated with males in a ratio of 3:1 at 25°C. All the crosses were carried out at 25°C unless mentioned otherwise (especially for few crosses where RNA was down regulated using UAS-RNAi lines).

6.8: UAS- GAL4 system

The UAS- GAL4 system, first described by Brand and Perrimon in 1993,. It is a very powerful technique for addressing gene function in vivo. In this bipartite approach, the expression of gene in question is under the control of UAS element, which responds to GAL4 protein. In the absence of this GAL4 the gene of interest remains transcriptionally silent. To activate the transcription the lines are mated to flies containing GAL4 expressing protein under the control of specific enhancers. Depending on the enhancers that control GAL4 expression, the gene of interest is transcribed in those regions (**Plate 6.5**).

6.8.1: GAL4 drivers used in the study

Omb-GAL4 (Calleja *et al.*,1996) :*Optomotor blind,* expresses in wing and haltere disc pouch

Ap-GAL4 (*et al*., 1995) :*Apterous,* expresses in dorsal cells of wing and haltere discs

En-GAL4(Brand and Perrimon, 1993) :*Engrailed,* expresses in the posterior compartment cells.

MS1096 GAL4 (Capidevila and Guerrero, 1994) : The driver expresses itself in the dorsal pouch region in the late third instar larva and during the pupal stage its expression extends to the entire pouch region.

Sal GAL4 : *Spalt,* the driver expresses itself along the A-P boundary of the wing blade.

Ptc GAL4(Brand and Perrimon, 1993) : *Patched*, the driver expresses itself along the A- P boundary in the anterior compartment of the wing imaginal disc.

6.8.2: Protein Trap line used in this study

DGRC #115100 : *Glec* protein trap line (w[1118];PBac{602.P.SVS-1}glec[CPTI000155] w[1118]; PBac{602.P.SVS-1}glec[CPTI000155])

6.8.3: UAS lines used in this study

UAS- *Glec* (generated in the study)

UAS- RNAi *Glec* (generated in the study)

6.9: Antibodies used in this study

Primary antibodies

Anti GFP (polyclonal from Invitrogen) : Used at a dilution of 1:5000

Anti Wingless (monoclonal from DSHB / Brook and Cohen, 1996): Used at a dilution of 1:500

Anti Delta (monoclonal from DSHB): Used at a dilution of 1:500

Anti Achaete(monoclonal from DSHB, Skeath and Carroll, 1991): Used at a dilution of 1:10

Anti Cut (monoclonal from DSHB, Blochlinger *et. al*., 1993): Used at a dilution of 1:20

Secondary antibodies

Goat Anti-mouse Alexa 488 (Molecular Probes, Invitrogen): Used at a dilution of 1:1000.

Goat Anti-mouse Alexa 568 (Molecular Probes, Invitrogen): Used at a dilution of 1:1000.

Goat Anti-rabbit Alexa 488 (Molecular Probes, Invitrogen): Used at a dilution of 1:1000.

Goat Anti-rabbit Alexa 568 (Molecular Probes, Invitrogen): Used at a dilution of 1:1000.

6.10: Generation of full length *Gliolectin* clone

Full length *Glec* mRNA was amplified from cDNA of *Drosophila* embryos. Embryos were collected and matured so that majority of them were at stage 13 (10 hours AEL) at the time of collection (it has been shown previously that *Glec* expresses maximally in embryos after stage 13). The embryos were flash frozen and then crushed in liquid nitrogen. The powder was then mixed with TriZol. The tissue was properly homogenised in TRIzol and the insoluble material was removed by centrifugation. The homogenised sample was incubated at room temperature for 5 minutes to permit complete dissociation of nucleoprotein complexes. 0.2 mL of chloroform was added to the mixture, shook vigoursly, incubated at room temperature for 10 minutes and then centrifuged to separate the phases. The upper aqueous phase was removed after

centrifugation and RNA was precipitated by adding 0.5 mL isoprapanol to it. The RNA precipitate was collected by centrifugation, washed with 70% ethanol and then dissolved in DEPC- treated water. The integrity of RNA was assessed by running an aliquot on an agarose gel while the purity was assessed by measuring its absorbance at 260, 280 and 230 nm.

Reverse Transcription was carried out using MMLV Reverse Transcriptase (Promega). Two micrograms of RNA was mixed with 0.5 µg of anchored dT primer and the mix was heated at 70° C for 5 minutes to denature the secondary structure of RNA. 5 µL of 5X Reverse Transcription buffer and 1.25 µL of dNTP was added and the mix was incubated at 25° C for 10 minutes. 1.5 µL of Reverse Transcriptase was added and incubated at 42° C for two hours. This cDNA was used to synthesize the full length *Glec*. The primers and the PCR program used for this purpose have been described in appendix

The amplified product was electrophoresed on an agarose gel and the band of required size was excised and eluted using QIAGEN gel elution kit. The amplicon was then tailed to generate poly A overhangs and cloned in pGEM-T easy vector. Conditions of adding A overhangs and cloning have been described in appendix. The cloned plasmid was used to transform DH5 alpha cells and the transformants were screened using colony PCR.

6.11: Preparation of *Glec* probe

In order to generate template for in vitro transcription, M13 forward and reverse primers were used to amplify the full length *Glec* from pGEM Teasy vector. The conditions of amplification have been described in appendix. RNA probe was prepared using the DNA template generated from the PCR using SP6 polymerase (Roche). The conditions of probe preparation have been described in appendix. The RNA was precipitated by adding 2.5 µl of 4 M LiCl and 75 µl of prechilled absolute ethanol to the reaction tube. The tube was kept at -70° C overnight. Following day the tube was centrifuged at 14000 rpm at 4°C for 30 minutes. The pellet was then washed

with 70% ethanol. The tube was re centrifuged and the supernatant was discarded. The resulting pellet was air dried and re suspended in DEPC water.

6.12: In situ hybridisation in imaginal discs of *Drosophila melanogaster*

The larvae were dissected and fixed in 4% formaldehyde at room temperature for 25 minutes with gentle rocking. The heads were then washed extensively in PBS. The heads were then dehydrated by passing in graded methanol (30%, 50%, 80%, 100 %) for 10 minutes each. The heads were then stored in methanol at -20° C. The heads were washed in a mix of xylene and ethanol(1:1) for 1 hour. Then they were washed in absolute ethanol for 10 minutes. They were then rehydrated by immersion in graded methanol. The heads were then treated with proteinase K (4 μg/mL) for 6 minutes at room temperature. The reaction was then stopped by adding glycine (2mg/mL) in PBS containing 0.2 % Tween 20 (PBT). The heads were then washed twice (5 minutes each in PBT). The heads were then re fixed in 4 % formaldehyde for 20 minutes. The heads were then washed five times (5 minutes each with PBT). The heads were then washed in hybridisation buffer :PBT (1:1) for 10 minutes at room temperature. The heads were then washed with hybridisation buffer for 10 minutes at room temperature.

The heads were then pre hybridised in hybridisation buffer at 70°C in hybridisation oven with gentle rocking for 2 hours. Hybridisation buffer with probe (1:500) was added and incubated at 57°C overnight. Before adding the probe to hybridisation buffer, the probe was heated to 99°C for 5 minutes together with salmon sperm DNA, chilled on ice and then added to preheated hybridisation buffer.

The hybridisation buffer was removed next day and the heads were washed with hybridisation buffer (minus probe) for 5 minutes, 10 minutes 30 minutes, 1 hour, 2 hour at 57°C. The heads were then washed with a mix of hybridisation buffer :PBT (1:1) at room temperature.

The heads were then washed with PBT and then blocked with 1 % blocking solution in PBT for 45 minutes at room temperature. Anti DIG- AP conjugated antibody (1:1000 in blocking buffer) was added to the heads and incubated for 2 hours at room temperature. The heads were again washed with PBT and then with AP buffer (20 minutes). NBT / BCIP solution (200 µl in 10 mL of AP buffer) was added . The colour reaction was stopped by adding PBS.

In order to perform fluorescent in situ hybridisation, one fast red tablet was dissolved in 2 mL of 0.1 M Tris Cl , pH 8.2. After the heads were washed after incubation with anti DIG antibody; they were incubated for 30 minutes in 0.1 M Tris Cl, pH 8.2. Fast red solution was added and the colour reaction was monitored under microscope. The reaction was stopped using PBS with 0.1 % triton 100 X.

6.13:Generation of Transgenics

In order to study the role of *Glec*, UAS – *Glec* (gain of function) and RNAi (loss of function) lines were generated.

6.13.1: Generation of UAS- *Glec* construct

In order to insert Glec clone in pUAST vector primers were synthesized, which had ECoR1 and Xho1 site in forward and reverse primer respectively. Full length clone of *Glec*was used as template and amplicon was obtained using the primers by PCR. The amplified product was then purified and restricted using ECoR1 and Xho1 and ligated in pUAST vector (insert:vector ration::3:1). The ligated product was used to transform DH5 alpha cells and the transformants were screened for the clones using colony PCR.

6.13.2: Construction of UAS-*Gliolectin*^{RNAi} construct

Two RNAi constructs were generated in SympUAST vector to target the same gene in two different regions.

The first RNAi construct spans7- 214 bp of the CDS of *Glec*.

The second RNAi construct spans 289- 682 bp of the CDS of *Glec*.

Both construct 1 and 2 were generated by amplifying the specific regions of the gene using primers containing specific restriction sites. Construct 1 had restriction sites for Xho1 and Not 1 in FP and RP respectively while construct 2 primers had restriction sites for ECoR1 and Xho1 in the FP and RP respectively. Both the constructs were amplified using the primers and restricted together with the vector (SympUAST) ; and were ligated with the restricted vector. The ligated plasmid was then used to transform DH5 alpha bacterial cells and the clones were screened using colony PCR. Positive clones were picked up, plasmid was isolated from them and used to generate transgenic flies after verification of clones.

6.13.3: Generation of transgenic flies using P elements

Transgenic flies were generated by injecting the modified pUAST / sympUAST vectors along with helper plasmid under low salt and slightly acidic conditions in embryos of w1118 flies during the first hour after egg laying (AEL).

6.13.4: Embryo Collection

Embryos of w1118 flies were collected by place a good number of freshly emerged flies in a fly cage with embryo collection medium (composition described in appendix). Fresh yeast paste was applied in the center of the embryo collection plate (size of a peanut) and the embryos were collected every hour in a plastic sieve.

6.14: Antibody staining of imaginal discs

Wandering late third instar fly larvae were collected in a chilled petri dish containing ice cold PBS (pH 7.4). The larval heads were cut and the gut material removed using a pair of forceps and scissors. The head were then flipped inside out and fixed in 4 % formaldehyde in PBS at room temperature for 20 minutes. The heads were then washed thrice with PBTX (0.1% Trition X- 100) at room temperature (each wash of 10 minutes). The heads were then blocked for 30 minutes at room temperature using

BSA (1% BSA in PBTX). Primary antibody was then used at the desired dilution and incubated with the heads overnight at 4ºC. The heads were then washed thrice with PBTX at room temperature (each wash of 10 minutes) and re blocked using 0.5 % BSA in PBTX for 30 minutes at room temperature. The heads were then incubated with secondary antibody at room temperature for 1 hour. The heads were washed again with PBTX, discs were dissected in PBS and mounted in 50 % glycerol.

6.15: Preparation of adult fly cuticle for mounting

The adult flies were first washed in 70% ethanol and then 10 % potassium hydroxide was added to cover the flies after removing the ethanol. The microfuge tube containing the fly was incubated at 70°C for 45 minutes and then the specimen was allowed to cool down to room temperature. Potassium hydroxide was cooled and enough water was added in the microfuge tube to cover the flies. The microfuge tube was again incubated at 70°C for 45 minutes. The specimen was allowed to cool to room temperature and then water was removed from the tube. 70% ethanol was added to the tube to wash the flies and later they were dehydrated using absolute ethanol. The flies were cleared using clove oil and mounted using DPX.

Results & Discussion

6.16:Expression of *Glec*

Expression pattern of *Glec* has already been shown in earlier studies by Prasad *et. al*.,2006. *Glec* is expressed along the D- V boundary in the wing imaginal discs and is not detectable in the haltere (based on in situ hybridisation studies). The protein trap lines showed expression of *Glec* in a more sensitive manner. *Glec* is expressed on the DV boundary (as also seen by in situ hybridisation) and is also seen in the intervein regions (between L3 and L4), flanked by Delta on either side. There is a general diffused pattern of Glec in the entire wing disc. However, the maximum expression is seen along the D- V boundary of wing discs. The expression of *Glec* on D V boundary was confirmed using double staining of *Cut* and anti GFP staining on protein trap line of *Glec* (**plates 6.6- 6.9**). In haltere discs, no specific expression pattern was observed in D/V boundary or presumptive veins. Low levels of diffused Glec all over the disc can be seen.

6.17: Generation of transgenic lines of *Glec*

Three independent transgenic lines were generated for each of the RNAi constructs- one targeting the N terminus of the *Glec* transcript while the other targeting C terminus of the mRNA. Three independent transgenic lines were also obtained for UAS-Glec which were used for ectopic expression of Glec using various GAL4 drivers. All the independent lines of each construct gave identical phenotypes, thus removing any possibilities of position effect or any other effect due to random insertion in the genome. For all the studies, homozygous viable lines for each construct were chosen- as they have the least possibility of disrupting any other gene.

6.18: Effectiveness of RNAi lines

In order to test the effectiveness of RNAi lines, GFP expression of *Glec* protein traps was analysed in lines where RNAi was driven by *Spalt*-GAL4 driver at 28°C. Single

copy of UAS-*Glec*^{RNAi} showed partial loss of Glec along the D-V boundary, while use of double copy of UAS-*Glec*^{RNAi} lines showed a near complete loss of Glec along the D-V boundary (**plate 6.10**).

6.19: Phenotypes obtained by down regulation of *Gliolectin*

Downregulation of *Glec* using various GAL4 drivers and UAS-*Glec*^{RNAi} gave identical phenotypes. In general, ectopic veins with bristles were obtained very close to the longitudinal veins. While ectopic veins had sharp boundaries as normal veins do, in few flies, broadening of wing veins was observed (**plates 6.11& 6.12**).

6.20: Phenotypes obtained by ectopic expression of *Gliolectin*

Ectopic expression of Glec resulted in loss of veins. Various GAL4 drivers, depending on the region where they express showed varying degree of loss of veins (**plate 6.13**).

6.21: Expression patterns of *Delta* and *Achaete* in wing imaginal discs when *Gliolectin* is down-or up-regulated

Downregulation of *Glec* in wing imaginal discs (using *MS1096*-GAL4 driver at 28°C) in a compartment specific manner led to a diffused expression of both *Delta* and *Achaete*(**plate 6.14**).

Ectopic expression of Glec in wing imaginal disc (using *MS1096*-GAL4 driver and *engrailed*-GAL4 driver at 28°C) led to down regulation of *Delta* expression in dorsal and posterior compartments, respectively (**plates 6.15 and 6.16**).

Gain and loss-of-function phenotypes observed for Glec appear to mimic gain and loss-of-function phenotypes of *Notch*. As already discussed, *Delta* expression in the third instar wing disc depends on the activity of EGFR pathway (Ve), which activates the transcription of *Delta* and other EGFR member genes. Delta protein, in

turn activates Notch in the neighbouring intervein cells and Notch in turn suppresses the expression of *Ve*, thus keeping its expression in the vein cells only.

The expression pattern of *Glec*, shows that it is expressed in a complementary manner to *Delta*. It is unlikely that *Glec* itself acts in the EGFR pathway to regulate the same negatively. It is more likely that it acts some where in the *Notch* signalling pathway to activate the *Notch* pathway and hence downregulate EGFR pathway indirectly.

If this hypothesis were true, then a gain of function of *Gliolectin* should activate Notch, which should reflect in its target genes, which needs to be explored further.

6.22: Effect of ectopic Gliolectin expression on wingless in halteres

In order to examine the effect of over-expression of Glec in haltere discs, *en*-GaL4 was used to ectopically express Glec in the entire posterior compartment of haltere. This resulted in ectopic activation of Wg expressing along the DV boundary in the compartment, which under normal circumstances is repressed in the posterior compartment. Analysis of *Delta* expression in these flies, showed that *Delta* levels were not elevated in the posterior compartment. It possible that Glec sensitises Notch so that it responds to Delta and Serrate even at very low concentrations.(**plate 6.17**).

As Glec plays a role in Notch signalling cascade and is sufficient to induce the expression of downstream targets of Notch in a context dependent manner; Ubx by bringing Glec under its control, exerts another level of control on Notch signalling pathway.

Summary

To summarise the work described in this chapter,:

1. *Glec* is expressed at the D-V boundary of wing imaginal discs and also along the prospective interveins. The use of protein trap lines showed its expression is maximal along the D-V boundary and the interveins and is less in the areas where Delta is expressed. Its expression is highly reduced or in haltere discs.
2. Loss of function of *Glec* showed phenotypes, which mimic loss of Notch function (or gain of EGFR activity). In contrast, gain of function of Glec showed phenotypes, which mimic gain of Notch (or loss of EGFR activity) function. It is, therefore, possible that Glec acts antagonistically to EGFR pathway.
3. Loss of function of Glec also downregulated target genes of Wg such as *Delta* and *Ach* in the wing imaginal disc, while gain of function of Glec induced the expression of Wg at the DV boundary of haltere in posterior compartment without affecting Delta expression.
4. By bringing Glec under its control Ubx exerts additional level of control over Notch signalling pathway.

Plates

Chapter 6

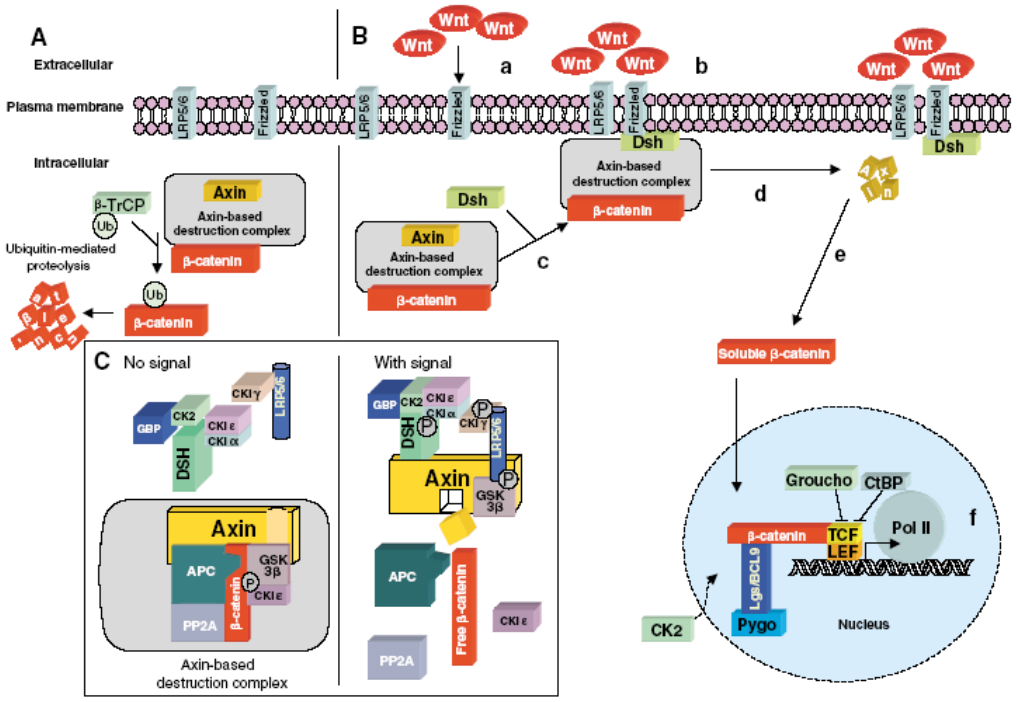


Plate 6.1: Core mechanism of Wnt/Wingless signalling mediated by β-catenin. (Image courtesy: Hayward *et.al.*, 2008)

A. In the absence of Wingless, a destruction complex forms around scaffolding protein, Axin which phosphorylates and subsequently destroys β-catenin.
B. In the presence of Wingless, hypophosphorylated form of β-catenin becomes available for entry into nucleus where it associates with factors like TCF to carry out transcription of various genes.

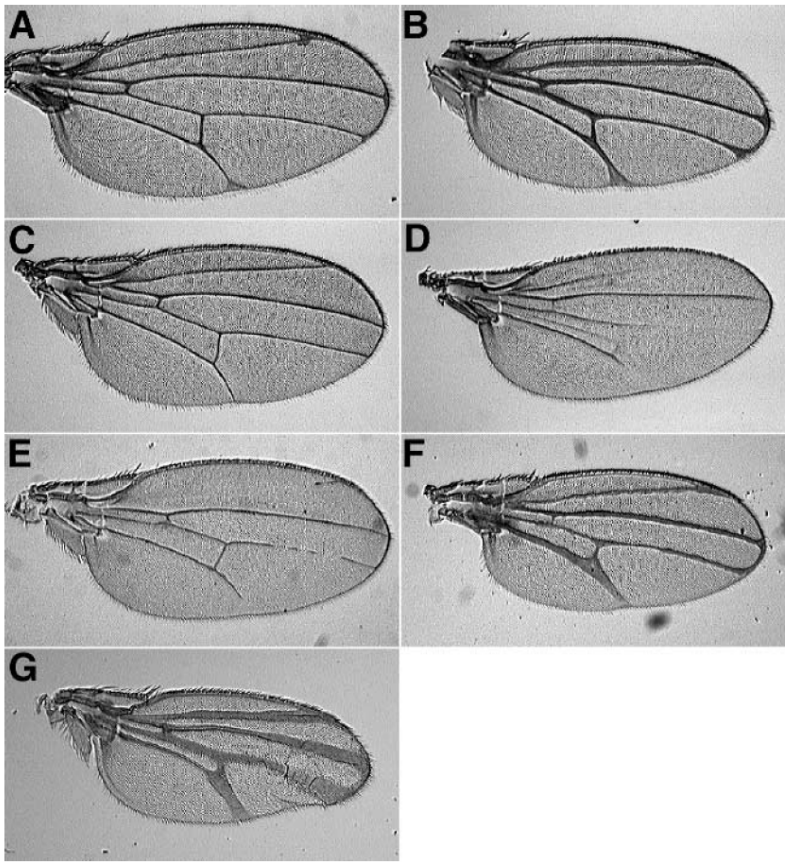


Plate 6.2: Phenotypic effects associated with perturbations in Delta- Notch signalling. (Image courtesy: Huppert *et. al.*, 1997)

(A,B) Adult wing phenotypes of Dl^{PlacZ}/Dl^{RF} animals reared at permissive temperature, 18°C (A), or pulsed to 32°C from 40 to 48 hours APF(B).

(C,D) Phenotypes of *P[hs-N(intra)]* animals grown at 25°C (C), or grown at 25°C and heat-pulsed to 37°C from 26 to 28 hours APF (D).

(E) Vein loss phenotype in a *UAS::DeltaWT/+; 1348::GAL4/+* wing from an animal grown at 25°C.

(F) Vein thickening phenotype in a *1348::GAL4/+; UAS::DeltaD/+* wing from an animal grown at 25°C.

(G) Extreme vein thickening phenotype of Dl^{EW}/Dl^{EW} *animals* grown at 25°C.

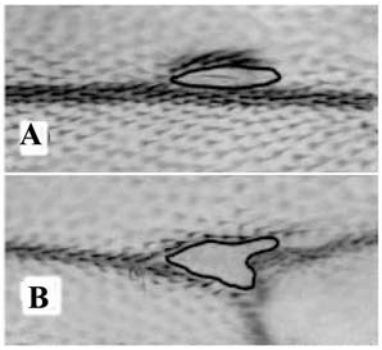


Plate 6.3: Examples of ectopic vein formation in Delta mutants (Image Courtsey : de Celis *et. al.*, 1997)

A. Dl ^{M1} clones in the dorsal L3
B. Dl ^{M1} clones in ventral L4

Loss of Delta induces vein differentiation in adjacent wild type cells provided they are very close to the vein cells.

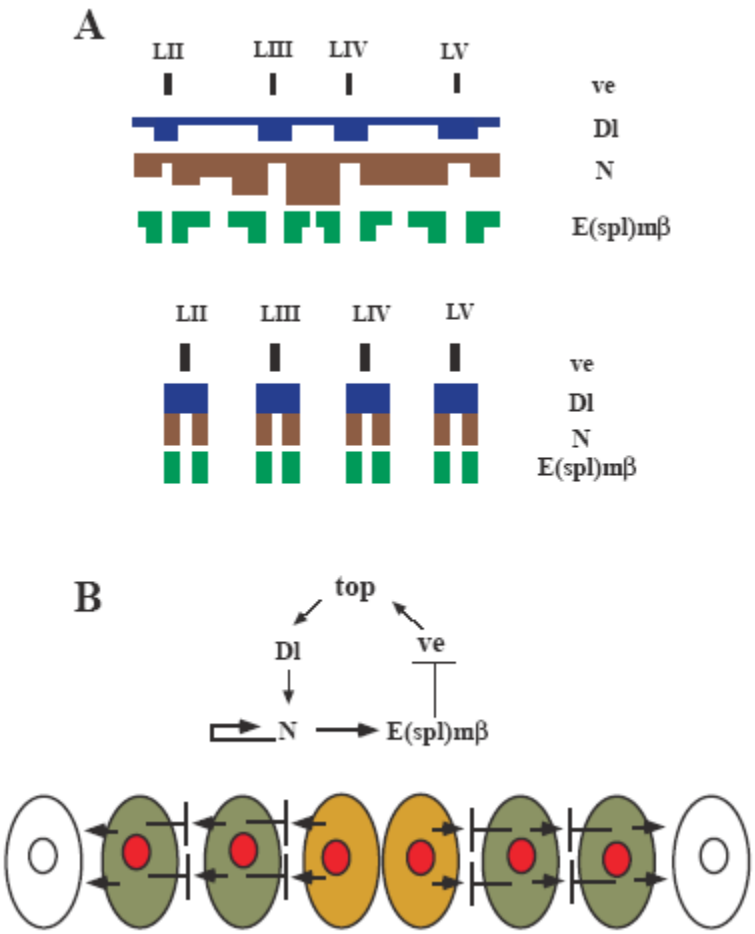


Plate 6.4: Genetic interactions that maintain the wing vein width (Courtsey: de Celis *et. al.*, 1997)

A: A diagrammatic representation of the wing disc showing the longitudinal veins L1- L4 in the top panel while the same veins are shown in the pupal disc in the bottom panel. The expression of *ve, Delta, Notch and E(Spl)mβ* has been shown. The height of the bars corresponding to each gene indicates their relative expression.

B: Summary of the genetic and cellular interactions in the maintenance of wing veins during pupal development.

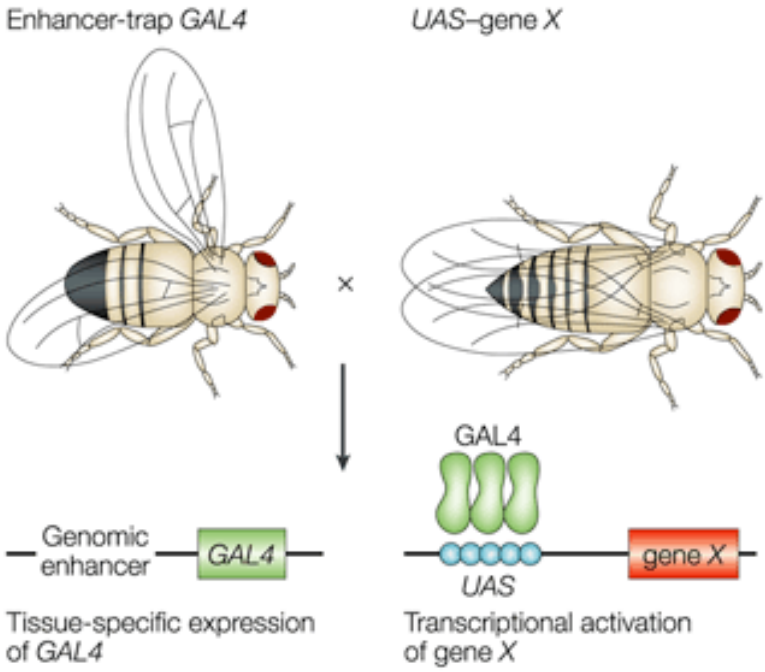


Plate 6.5: UAS GAL4 system in *Drosophila melanogaster* (Nature Reviews)

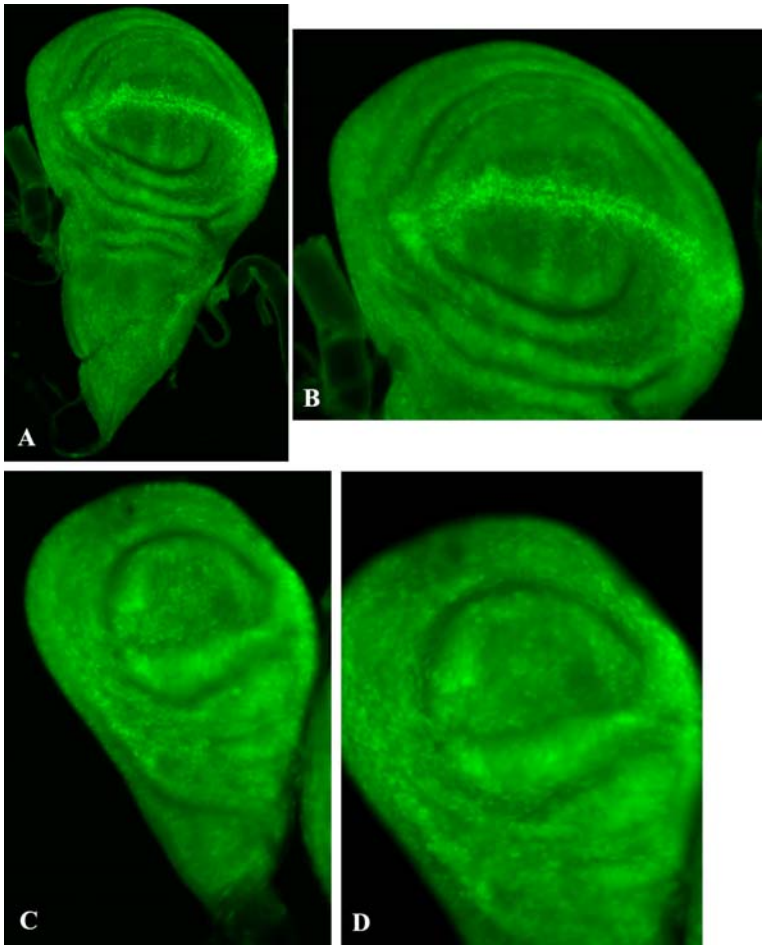


Plate 6.6: Expression pattern of Glec as revealed using anti GFP antibody staining on a Glec protein trap line (DGRC #115100) shown in green.

A. Glec expression in wing imaginal disc of late third instar larva at lower magnification.
B. Glec expression in wing imaginal disc of late third instar larva at higher magnification.
C. Glec expression in haltere imaginal disc of late third instar larva at lower magnification.
D. Glec expression in haltere imaginal disc of late third instar larva at higher magnification.

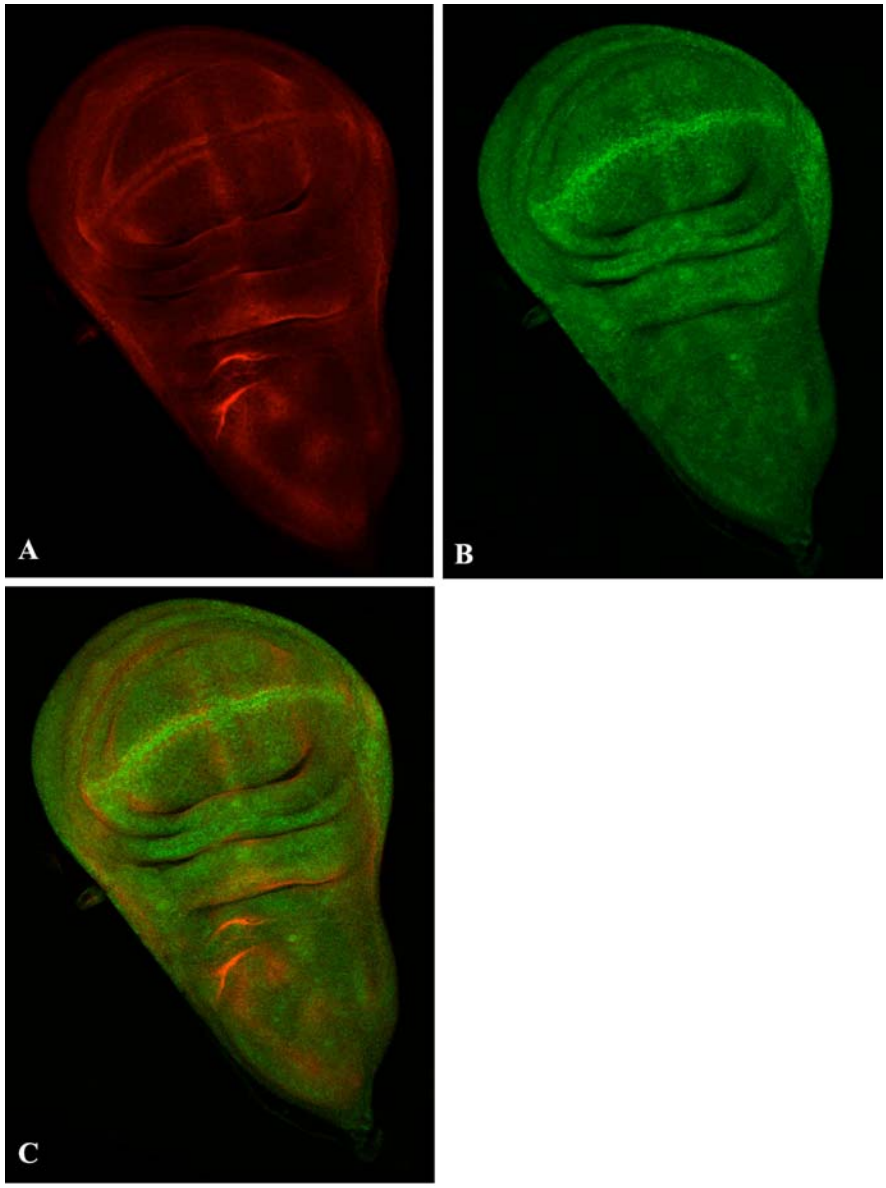


Plate 6.7: Complementary expression of Glec and Delta in the third instar wing imaginal disc of *Drosophila melanogaster*.

A. Expression of Delta in wing imaginal disc of late third instar larva (shown in red).
B. Expression of Glec in wing imaginal disc of late third instar larva (shown in green).
C. Merged image showing the complementary expression of Glec and Delta in the wing imaginal disc of late third instar larva.

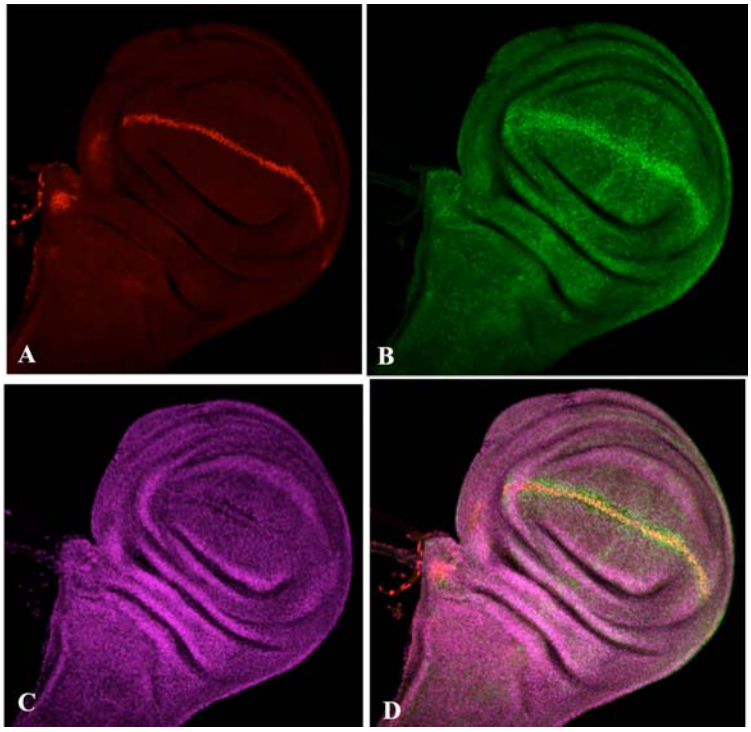


Plate 6.8: Localisation of Glec with respect to Ct in the third instar wing imaginal disc of *Drosophila melanogaster*.

A. Expression of Ct in wing imaginal disc of late third instar larva (shown in red).
B. Expression of Glec in wing imaginal disc of late third instar larva (shown in green).
C. Nuclei of wing imaginal disc stained with DAPI (shown in purple).
D. Merged image showing co- localisation of Glec and Ct at the D-V boundary.

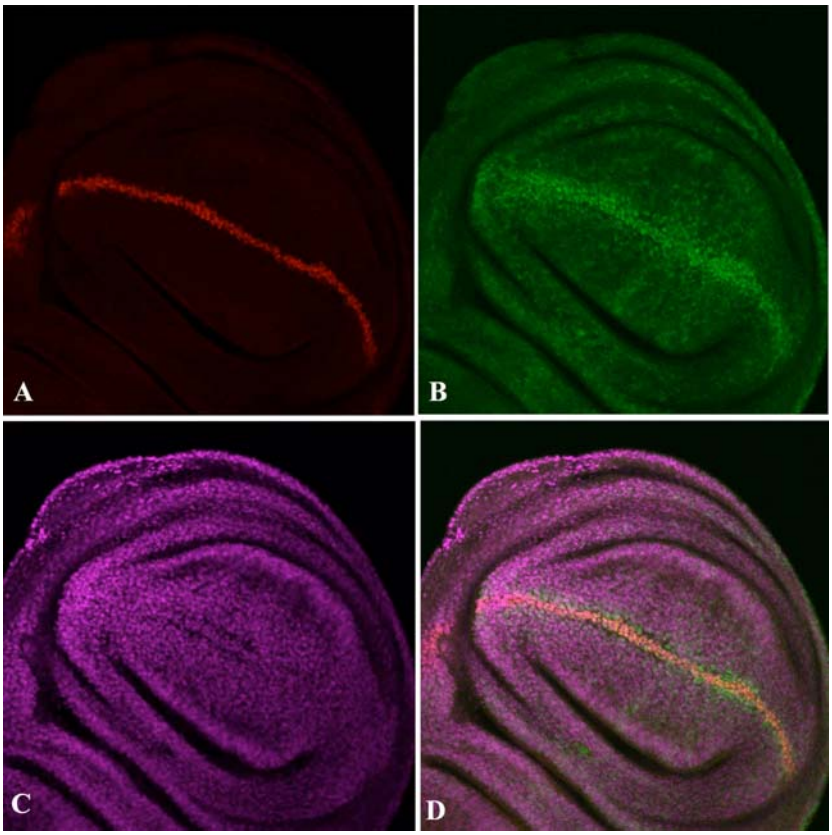


Plate 6.9 : Localisation of Glec with respect to Ct in the third instar wing imaginal disc of *Drosophila melanogaster* (higher magnification).

A. Expression of Ct in wing imaginal disc of late third instar larva (shown in red).
B. Expression of Glec in wing imaginal disc of late third instar larva (shown in green).
C. Nuclei of wing imaginal disc stained with DAPI (shown in purple).
D. Merged image showing co- localisation of Gliolectin and Cut at the D-V boundary.

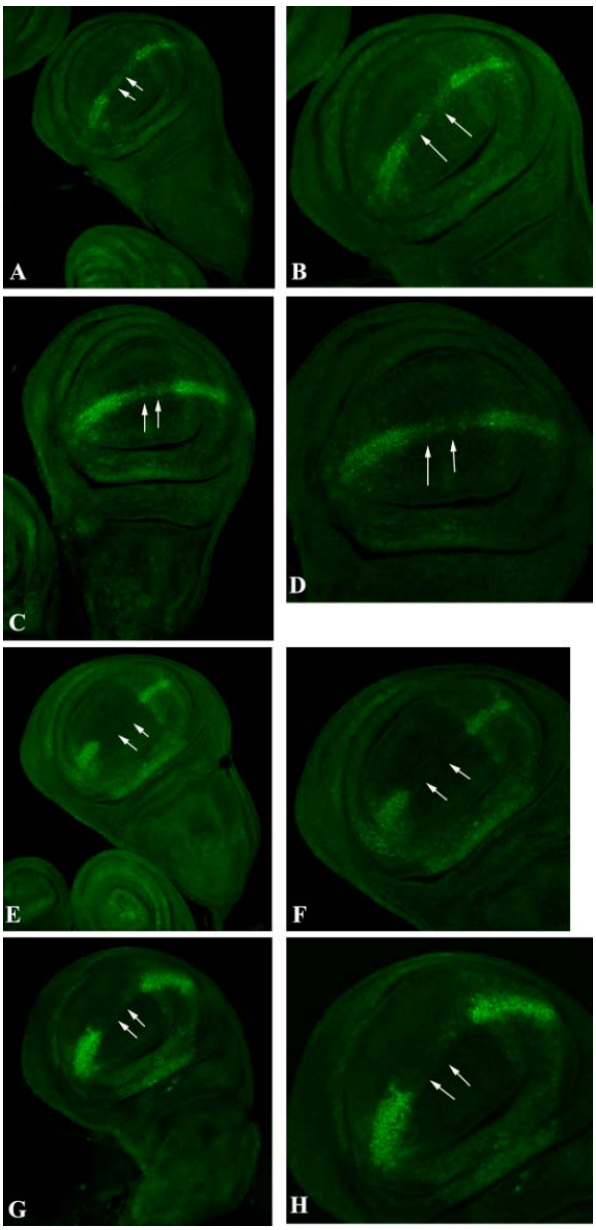


Plate 6.10: Effect of various RNAi constructs on Glec expression at 28ºC (glec expression has been shown in green).

A,B : Downregulation of Glec as detected in protein trap line when *Sal* GAL 4 driver was used to drive expression of UAS *Glec*^{RNAi} line (8-4) at 28ºC. (*Sal* Gal4/ UAS *Glec*^{RNAi} 8-4; DGRC#115100/+).

C,D : Downregulation of Glec as detected in protein trap line when *Sal* GAL 4 driver was used to drive expression of UAS *Glec*^{RNAi} line (19-3) at 28ºC. (*Sal* Gal4/+; DGRC#115100/UAS *Glec*^{RNAi} 19-3).

E,F : Downregulation of Glec as detected in protein trap line when *Sal* GAL 4 driver was used to drive expression of UAS RNAi line (10-1) at 28ºC. (*Sal* Gal4/ UAS *Glec*^{RNAi} 10-1; DGRC#115100/+)

G,H : Downregulation of Glec as detected in protein trap line when *Sal* GAL 4 driver was used to drive expression of 2 copies of UAS *Glec*^{RNAi} line (8-4; 19-3) at 28ºC. (*Sal* Gal4/ UAS *Glec*^{RNAi} 8-4; DGRC#115100/UAS *Glec*^{RNAi} 19-3)

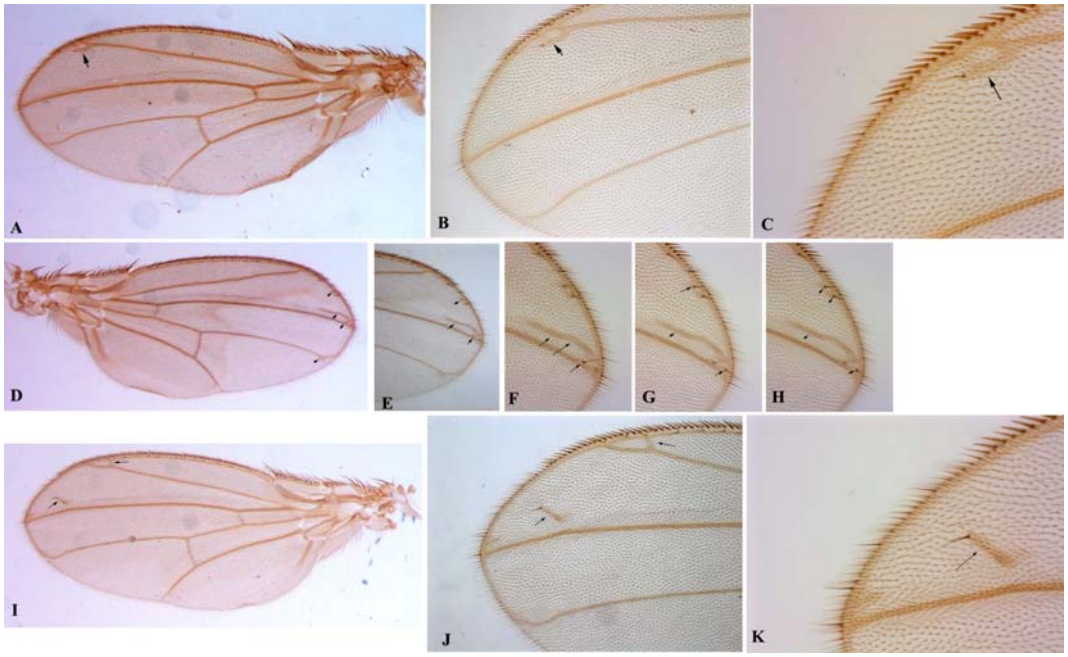


Plate 6.11: Effect of loss of *Glec* on the wing blade using MS1096 GAL4 driver and various RNAi lines at 28ºC.

A,B,C: Effect of down regulation of *Glec* using UAS *Glec*^{RNAi} (3-1) construct 1, targeting the N- terminus of *Glec* transcript. Ectopic vein, attached to the L2 vein and a bristle is observed. Genotype : MS1096GAL4; UAS *Glec*^{RNAi} (3-1).

D,E, F, G, H: Effect of down regulation of *Glec* using UAS *Glec*^{RNAi} (8-4) construct 2, targeting the C- terminus of *Glec* transcript. Ectopic vein, attached to the L3 vein and numerous bristle are observed. Genotype : MS1096GAL4; UAS *Glec*^{RNAi} (8-4).

I,J,K : Effect of down regulation of *Glec* using UAS *Glec*^{RNAi} (10-1) construct 2, targeting the C- terminus of *Glec* transcript. Ectopic vein, close to the L3 vein and a bristle is observed. Genotype : MS1096GAL4; UAS *Glec*^{RNAi} (10-1).

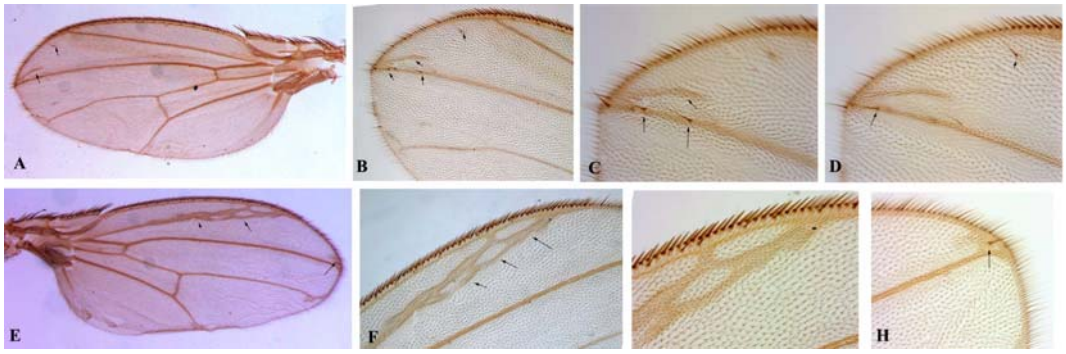


Plate 6.12: Effect of loss of *Glec* on wing blade (using MS1096 GAL4 and Sal GAL4 driver and different RNAi lines)

A,B,C,D: Effect of down regulation of *Glec* using UAS *Glec*^{RNAi} (19-3) construct 2, driven by MS1096 GAL4 at 28ºC targeting the C- terminus of *Glec* transcript. Ectopic vein, close to the L3 vein and bristles are observed. Genotype : MS1096GAL4; UAS *Glec*^{RNAi} (19-3).

E,F,G,H: Effect of down regulation of *Glec* using UAS double RNAi (8-4; 19-3), driven by *Sal* GAL4 at 28ºC targeting the C- terminus of *Glec* transcript. Ectopic vein, close to the L3 vein and a bristle is observed. Also the L2 vein shows broadening. Genotype : *Sal* GAL4; UAS *Glec*^{RNAi} (8-4); UAS *Glec*^{RNAi} (19-3).

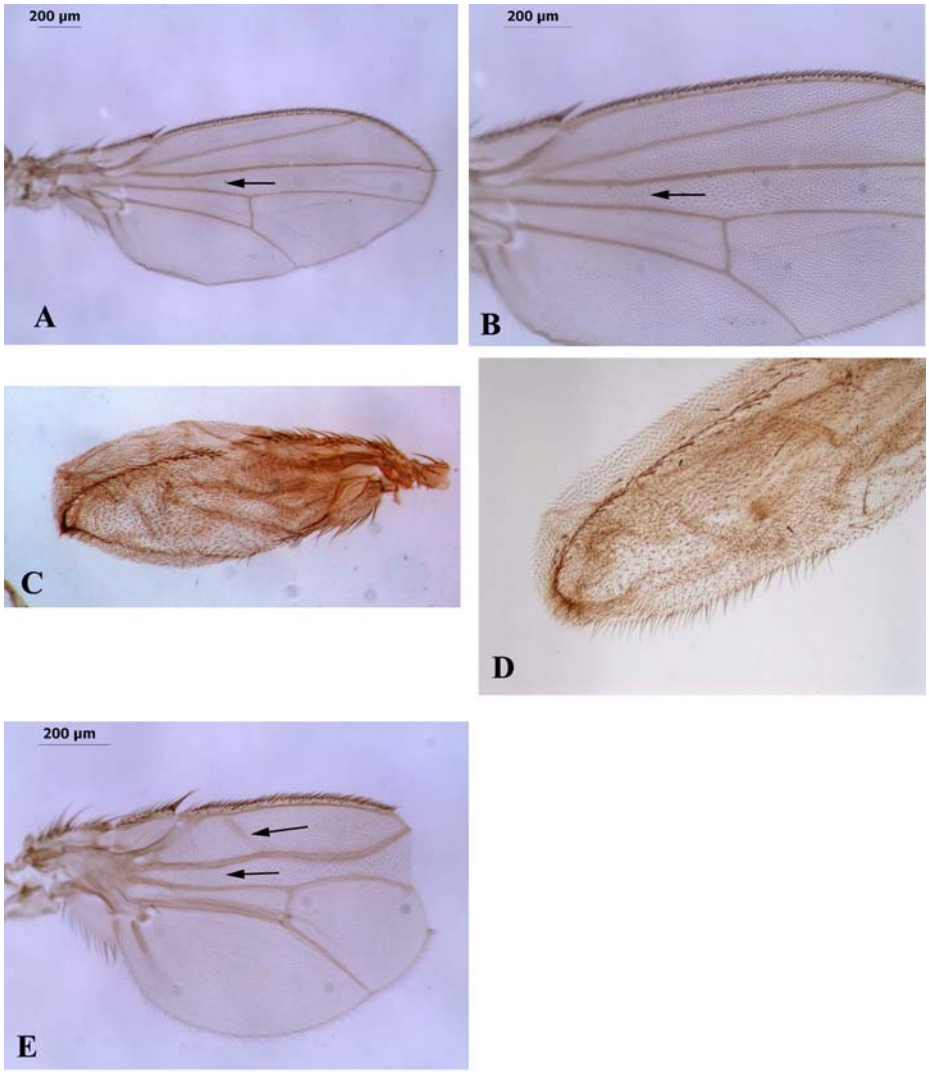


Plate 6.13: Effect of ectopic expression of *Glec* on wing blade of *Drosophila melanogaster* (using various GAL4 drivers and UAS *Glec* (2-10) at 28ºC)

A,B: Effect of ectopic expression of *Glec* on wing blade of *Drosophila melanogaster* using *Ptc* GAL4 at 28ºC. (Genotype: *Ptc* GAL4; UAS *Glec*). Loss of Cross vein1 is clearly visible.

C,D: Effect of ectopic expression of *Glec* on wing blade of *Drosophila melanogaster* using MS1096 GAL4 at 28ºC. (Genotype: MS1096 GAL4; UAS *Glec*). All the veins in the wing blade are lost.

E: Effect of ectopic expression of *Glec* on wing blade of *Drosophila melanogaster* using *Omb* GAL4 at 28ºC. (Genotype: *Omb* GAL4; UAS *Glec*). Loss of longitudinal vein 2 can be seen.

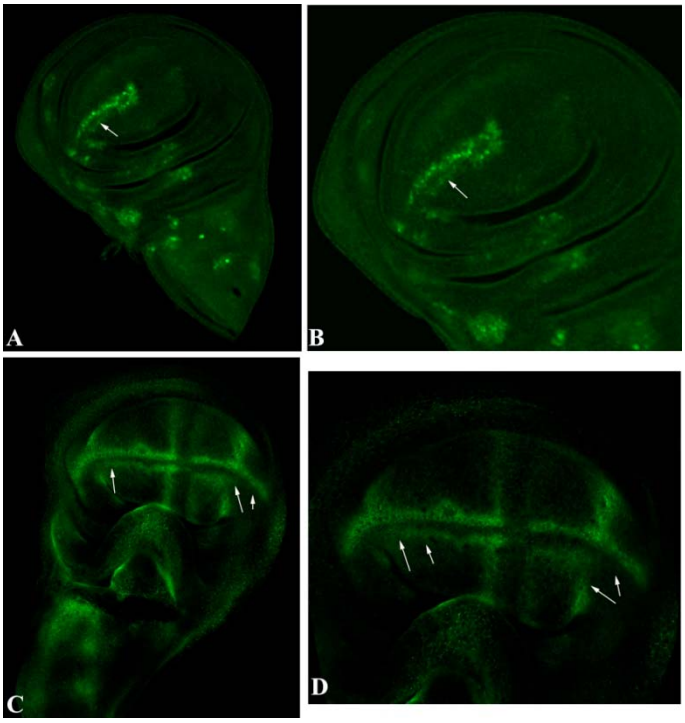


Plate 6.14: Effect of loss of *Glec* on Delta and Ach in wing imaginal discs (down regulation of *Glec* was achieved by using MS1096 GAL4 driver and two copies of UAS *Glec*^{RNAi} at 28ºC). Genotype : MS1096 GAL4; UAS *Glec*^{RNAi} 8-4; UAS *Glec*^{RNAi} 19-3).

A,B: Ach expression when *Glec* is down regulated in the dorsal compartment of wing imaginal discs. The expression in the dorsal compartment is diffused (shown in green).

C,D: Delta expression when *Glec* is down regulated in the dorsal compartment of wing imaginal discs. The expression in the dorsal compartment is diffused and seems to be lower than that in the ventral compartment (shown in green).

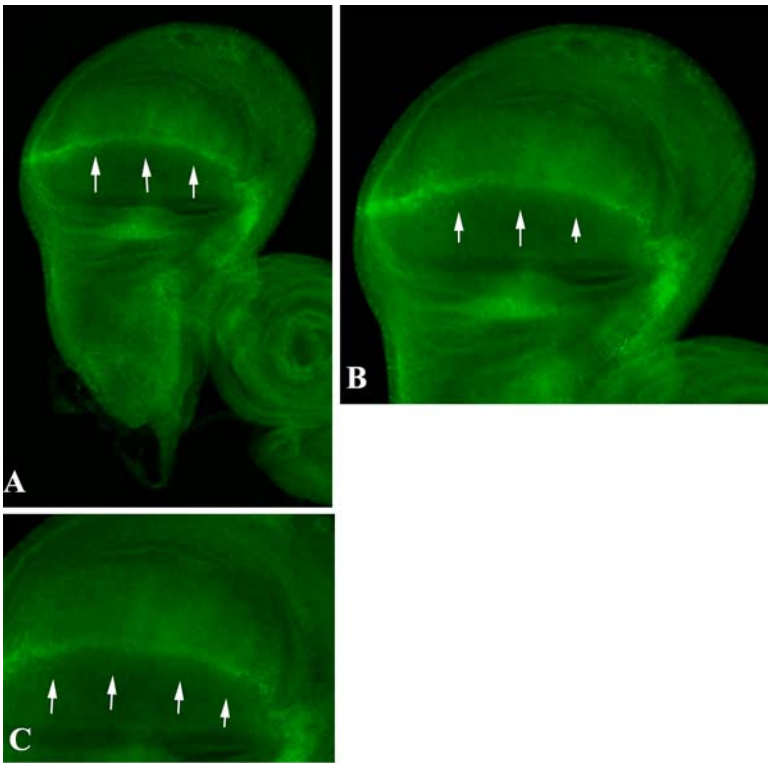


Plate 6.15: Effect of ectopic expression of *Glec* on Delta expression. (shown in green) (*Glec* was ectopically expressed in the dorsal compartment of wing imaginal discs using MS1096 GAL4 at 28ºC). Genotype : MS1096 GAL4; UAS *Glec* (2-10).

A,B,C: Delta expression in wing imaginal discs of late third instar larva when *Glec* is ectopically expressed in the dorsal compartment of wing discs. Delta expression in the dorsal compartment completely goes down on ectopic expression of *Glec*.

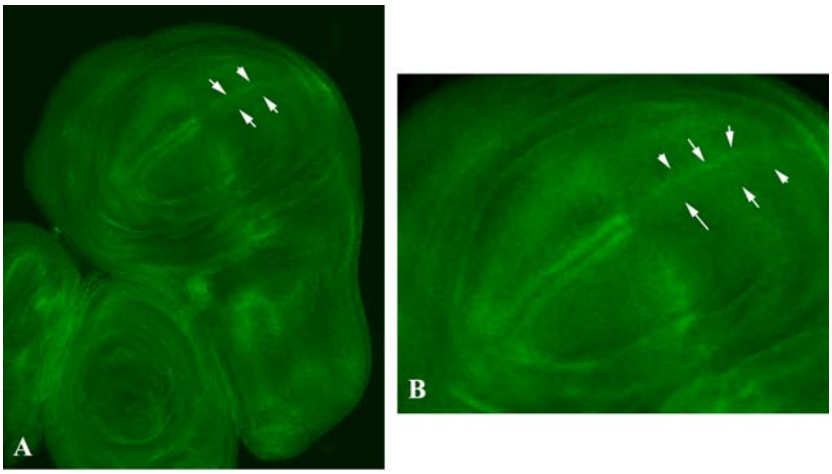


Plate 6.16: Effect of ectopic expression of *Glec* on Delta expression. (shown in green). (*Glec* was ectopically expressed in the dorsal compartment of wing imaginal discs using *en* GAL4 at 28ºC). Genotype: *engrailed* GAL4; UAS *Glec* (2-10).

A,B,C: Delta expression in wing imaginal discs of late third instar larva when *Glec* is ectopically expressed in the posterior compartment of wing discs. Delta expression in the posterior compartment goes down on ectopic expression of *Glec*.

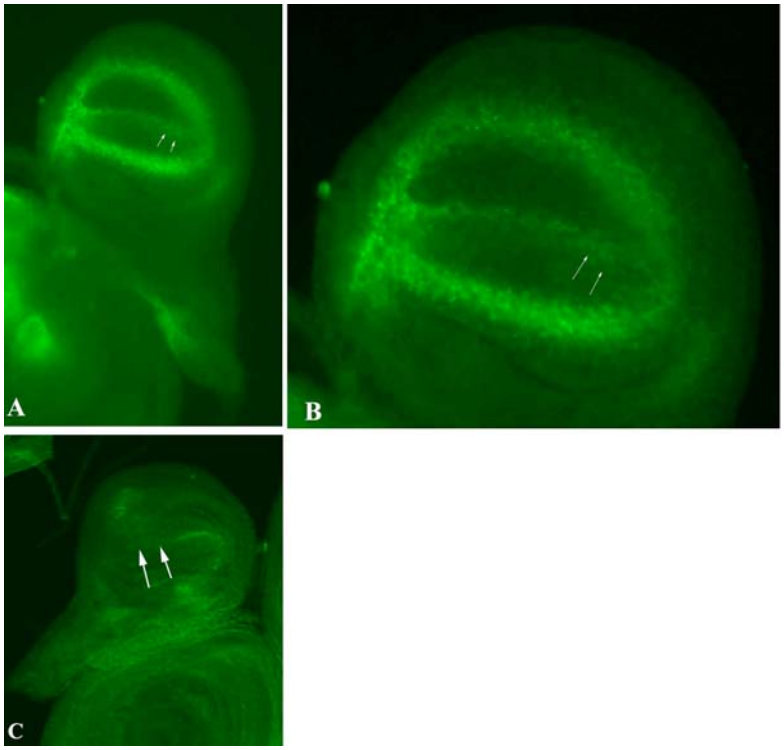


Plate 6.17: Effect of ectopic expression of *Glec* in the posterior compartment of haltere imaginal disc on Wingless and Delta expression at 28 ºC. Genotype : *en* GAL4; UAS *Glec*

A, B: Wingless expression in the flies of above genotype. Wingless expression is seen in the posterior compartment where usually it is suppressed. (shown in green).

C: Delta expression in the haltere imaginal disc of flies of above genotype. Delta expression is not enhanced in the posterior compartment(shown in green).

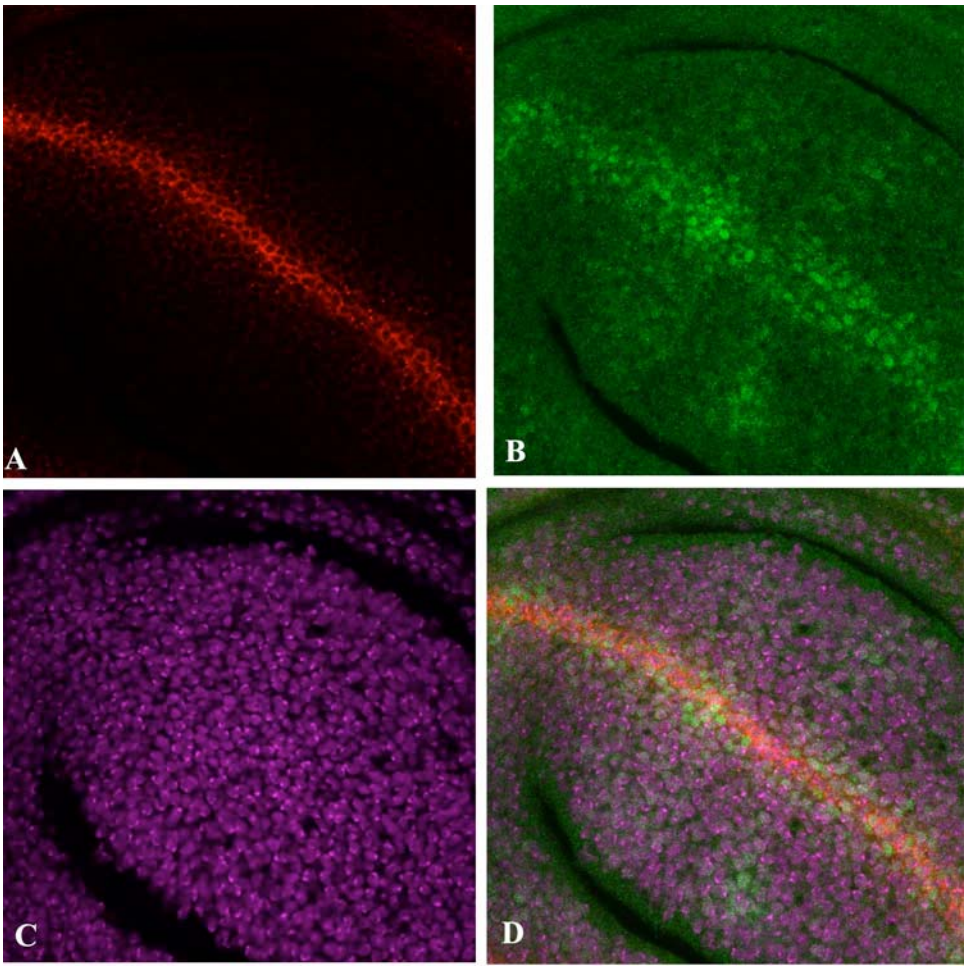


Plate 6.18: Localisation of Gliolectin in cells of *Drosophila melanogaster*

A. Wg expression along the D-V boundary of wing imaginal discs (Wingless is associated with cell membranes, shown in red)
B. Glec expression in the wing pouch of wing imaginal discs (Higher expression of Glec is seen along the D-V boundary and the interveins, shown in green)
C. Nuclei in the wing imaginal discs detected by DAPI staining (shown in purple)
D. A merged image of all the three showing that Glec is neither associated with cell membrane nor with nucleus.

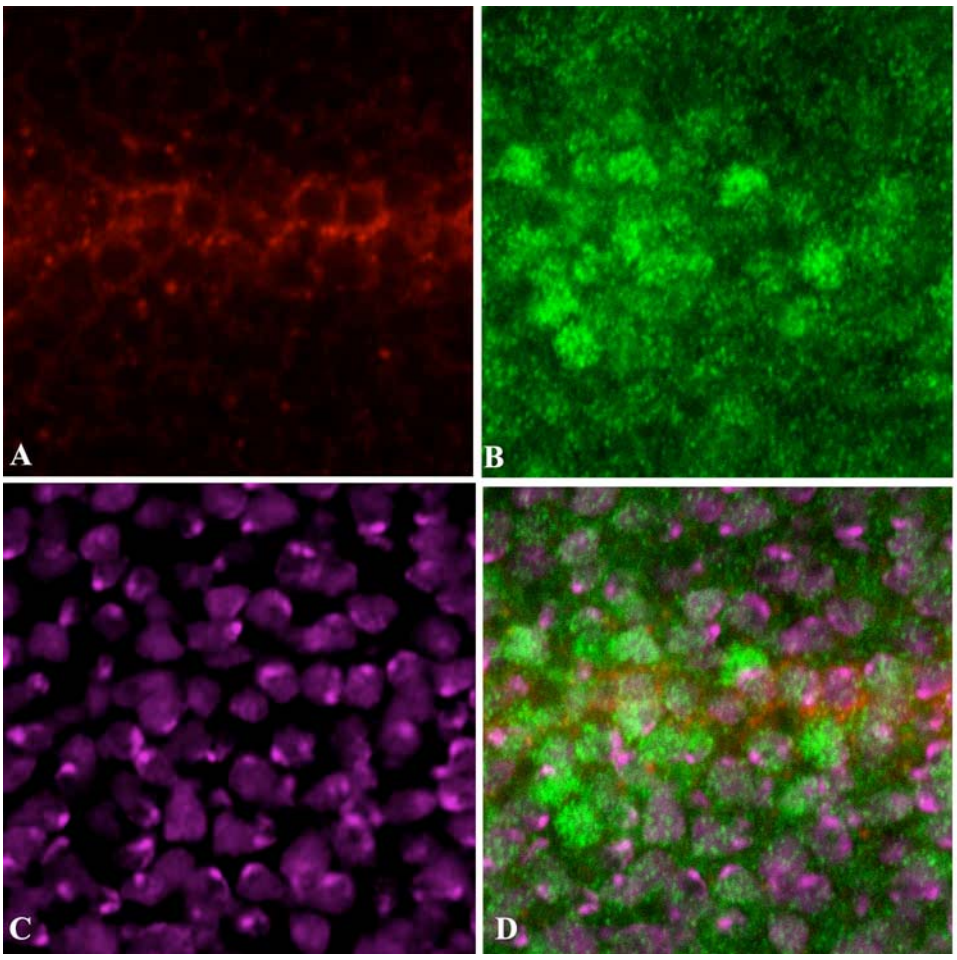


Plate 6.19: Localisation of Glec in cells of *Drosophila melanogaster*

A. Wg expression along the D-V boundary of wing imaginal discs (Wingless is associated primarily with cell membranes) (shown in red).
B. Glec expression in the wing pouch of wing imaginal discs (Gliolectin is seen as punctate dots in the cells of *Drosophila melanogaster*.) (shown in green).
C. Nuclei in the wing imaginal discs detected by DAPI staining (shown in purple).

D. A merged image of all the three showing that Glec is neither associated with cell membrane nor with nucleus.

Future work

Work carried out in this study showed that while Ubx has acquired a number of novel target genes in dipterans to shape a haltere instead of hind wing in the early hymnopterans, the protein has also regulated a significant number of its pre existing targets in a different way in the dipterans. The expression pattern of few genes and bioinformatics analysis of their promoters clearly showed the evolution of binding sites of new cofactors in their promoter region bound by Ubx which might have brought a change in regulation of these genes in dipteran haltere when compared to the hymenopteran hind wing.

A detailed analysis of these promoters (bound by Ubx in both- the hind wing disc of honey bee and in the haltere disc of fruit fly) using reporter assays need to be performed to exactly understand the evolution of which motifs have caused a change in regulation of Ubx target genes (Ubx being bound to the promoters of these genes in both the orders). This involves cloning of Ubx bound promoter regions in honey bee in suitable expression vectors and studying their expression in the wing and haltere of fruit fly. Also, suitable motifs might be inserted or deleted from this honey bee promoter sequence (to mimic the evolved fruit fly promoter sequence), without altering the Ubx binding sites and the expression of the modified promoter may again be studied in the fruit fly. Such analysis will help us to identify the motifs that have either evolved in the dipteran lineage or have been lost from hymenopteran lineage to regulate the pre existing targets of Ubx in a novel way in the dipterans.

As the diptera specific target genes of Ubx must have played an important role in shaping the haltere, characterization of few important diptera specific target genes of Ubx should be carried out to understand their role in the development of wing and haltere of fruit fly. Such analysis would help us understand the biological significance of the genes that have been brought under the control of Ubx in dipteans and how their regulation have helped shape the haltere instead of a wing.

Tables generated in this study

Replicate 1 samples	Total number of reads obtained	Number of reads that mapped to unique locations	Number of reads that mapped to multiple locations	Number of reads that failed to align
FW Input	29017278	22631097 (77.99%)	3077354 (10.61%)	3308827 (11.40%)
FW pre immune	36562872	8386960 (22.94%)	1466258 (4.01%)	26709654 (73.05%)
FW anti Ubx	41626439	13858852 (33.29%)	2169938 (5.21%)	25597649 (61.49%)
HW Input	28154578	21934571 (77.91%)	3185174 (11.31%)	3034833 (10.78%)
HW pre immune	28361093	7720127 (27.22%)	1289297 (4.55%)	19351669 (68.23%)
HW anti Ubx	35691147	14561757 (40.80%)	2051507 (5.75%)	19077883 (53.45%)

Table 1: Number of reads obtained from replicate 1 (hive 10) in various chipped DNA samples

Replicate 2 samples	Total number of reads obtained	Number of reads that mapped to unique locations	Number of reads that mapped to multiple locations	Number of reads that failed to align
FW Input	22340337	17298473 (77.43%)	2600796 (11.64%)	2441068 (10.93%)
FW pre immune	18820978	11794124 (62.66%)	1766975 (9.39%)	5259879 (27.95%)
FW anti Ubx	22174088	9112663 (41.10%)	1456114 (6.57%)	11605311 (52.34%)
HW Input	20212052	15994236 (79.13%)	2254751 (11.16%)	1963065 (9.71%)
HW pre immune	19539102	10581141 (54.15%)	1468683 (7.52%)	7489278 (38.33%)
HW anti Ubx	22756942	17195379 (75.56%)	1695785 (7.45%)	3865778 (16.99%)

Table 2: Number of reads obtained from replicate 2 (hive 8) in various chipped DNA samples

Dataset	Ubx targets in hind wing disc of *Apis* (current study)	Ubx targets in haltere discs of *Drosophila* (Choo *et. al*., 2011)	Ubx targets in haltere discs of *Drosophila* (Agrawal *et. al*., 2011)
Ubx targets in hind wing disc of *Apis* (current study)	1182	195	88
Ubx targets in haltere discs of *Drosophila* (Choo *et. al*., 2011)	195	1170	219
Ubx targets in haltere discs of *Drosophila* (Agrawal *et. al*., 2011)	88	219	557

Table 3: Comparison of Ubx targets in hind wing of honeybee and those in halteres of fruit fly

Flybase Gene Id	**Associated Biological processes**	**Common name of the gene**
FBgn0000179	It is involved in anterior posterior pattern formation of the wing disc (inferred from mutant phenotype). It is also involved in organ growth.	*bifid*
FBgn0000352	The gene is involved in positive regulation of *hh* target transcription factor activity (inferred from mutant phenotype).	*costa*
FBgn0000448	The gene has been known to regulate development and metamorphosis.	*hormone receptor-like*
FBgn0000463	The gene is responsible for activation of Notch, is responsible for apposition of dorsal and venral imaginal disc derived wing surfaces. It also plays an important role in lateral inhibiton.	*Delta*
FBgn0000492	The gene is involved in dorsal/ventral pattern formation in the wing disc. It is also involved in wing and notum subfield formation.	*Drop*
FBgn0000497	The gene is involved in cell proliferation and wing disc pattern formation. It is also involved in establishment of imaginal disc-derived wing hair	*dachsous*

	orientation.	
FBgn0000542	It is involved in ubiquitin-dependent protein catabolic process.	*echinus*
FBgn0000546	The gene is involved in wing morphogenesis and cell adhesion (genetically interacts with *if* and *mys*).	*ecdysone receptor*
FBgn0000567	The gene is involved in apoptosis.	*ecdysone-induced protein 74EF*
FBgn0000568	The gene is involved in regulation of gene expression.	*ecdysone-induced protein 75B*
FBgn0000577	The gene is involed in anterior- posterior pattern formation of the wing disc.	*engrailed*
FBgn0000721	The gene is involved in protein phosphorylation and affects larval locomotory behaviour.	*foraging*
FBgn0001075	The gene is involed in establishment of imaginal disc-derived wing hair orientation It is negative regulator of growth and Wnt receptor signaling pathway . It is also involved in Hippo signaling cascade (inferred from genetic interaction with *wts*).	*fat*

FBgn0001083	The gene is involved in imaginal disc-derived wing morphogenesis.	*furrowed*
FBgn0001138	It is involved in organ morphogenesis and tissue development.	*grain*
FBgn0001228	response to heat	*heat shock gene 67Bb*
FBgn0001253	The gene is involed in imaginal disc eversion.	*ecdysone-inducible gene E1*
FBgn0001269	The gene is involved in wing disc anterior/posterior pattern formation.	*invected*
FBgn0001297	The gene is involved in wing disc development, imaginal disc fusion, thorax closure . The gene is also involved in regulation of cyclin-dependent protein serine/threonine kinase activity involved in G2/M.	*kayak*
FBgn0001332	The gene is involed in negative regulation of Wnt receptor signalling pathway.	*Lobe*
FBgn0001994	-	*cropped*
FBgn0002413	The gene is involved in imaginal disc growth, establishment of imaginal disc-derived wing hair	*discs overgrown*

	orientation. The gene is also involved in Wnt receptor signaling pathway.	
FBgn0002643	The gene is involved in apposition of dorsal and ventral imaginal disc-derived wing surfaces, wing margin morphogenesis and asymmetric cell division. The gene is involved in Notch signaling pathway.	*mastermind*
FBgn0002948	meiotic chromosome segregation | inferred from mutant phenotype	*no distributive disjunction*
FBgn0003016		*outspread*
FBgn0003028	The gene is involved in non- sensory hair organization.	*ovo*
FBgn0003079	The gene is involved in wing and notum subfield formation . The gene is also involved in epidermal growth factor receptor signalling pathway.	*pole hole*
FBgn0003118	The gene is involved in wing disc dorsal/ventral pattern formation (inferred from genetic interaction with *Bx*, *ap*) It is also involved in epidermal growth factor receptor signaling pathway	*pointed*

	.	
FBgn0003138	The gene is involved in negative regulation of JAK-STAT cascade.	*protein tyrosine phosphatase 61F*
FBgn0003165	The gene negatively regulates epidermal growth factor receptor signalling pathway.	*pumilio*
FBgn0003175	The gene is involved in imaginal disc-derived wing vein morphogenesis.	*plexus*
FBgn0003277	It serves as a DNA damage checkpoint protein at G2 stage of mitosis.	*RNA polymerase II 215kD subunit*
FBgn0003319	The gene is involved in morphogenesis of larval imaginal disc epithelium.	*Stubble*
FBgn0003339	The gene is involved in segment specification and sex comb development.	*Sex combs reduced*
FBgn0003415	The gene is involved in wing disc dorsal/ventral pattern formation (inferred from genetic interaction with *Bx*, *ap*). It is also a positive regulator of Wg signalling pathway.	*skuld*
FBgn0003499	The gene is involved in ectoderm development.	*stripe*

FBgn0003502	The gene is involved in JNK cascade, affects dorsal closure and is involved in wing morphogenesis.	*Btk family kinase at 29A*
FBgn0003651	The gene is involved in neuron-neuron synaptic transmission, ventral cord and fat body development.	*seven up*
FBgn0003716	The gene is involved in Decapentaplegic signalling pathway. It is involved in dorsal closure and cell elongation during wing disc morphogenesis.	*thickveins*
FBgn0003731	The gene is involved in proximal/distal pattern formation in the wing disc. It is also involved in wing vein specification.	*epidermal growth factor receptor*
FBgn0003744	The gene is involved in organization of wing hair.	*tricornered*
FBgn0003944	The gene is involved in specification of segmental identity and responsible for haltere development.	*Ultrabithorax*
FBgn0003963	The gene is associated with torso signalling pathway and is involved in dorsal closure.	*u-shaped*
FBgn0003975	The gene is responsible for development of wing blade and wing margin.	*vestigial*
FBgn0004009	The gene is involved in wing cell fate specification and dorsal/ventral pattern formation of the wing imaginal disc.	*wingless*

FBgn0004197	The gene is involved in notch signalling pathway and is responsible for wing morphogenesis. It is responsible for imparting asymmetry in signalling along D-V axis.	*Serrate*
FBgn0004394	The gene is involved in neuroblast development .	*POU domain protein 2*
FBgn0004462	The gene product is responsible for protein phosphorylation.	*Protein kinase-like 17E*
FBgn0004583	The gene negatively regulates imaginal disc growth. It also shows interaction with dpp , showing it is involved in Dpp signalling pathway. Genetic interaction of the gene with wts and physical association with yki associate it with hippo signalling pathway.	*expanded*
FBgn0004606	The gene is involved in negative regulation of transcription from RNA polymerase II promoter.	*Zn finger homeodomain 1*
FBgn0004647	The gene is involved in pattern formation of the wing disc along the dorsal- ventral axis.	*Notch*
FBgn0004655	The gene product is involved in segregation of chromosomes during	*wings apart-like*

FBgn0004858	The gene is involved in Notch signalling pathway.	*elbow B*
FBgn0004860	The gene is responsible for chromatin silencing.	*polyhomeotic distal*
FBgn0004861	The gene product is responsible for silencing various genes (inferred from genetic interaction with Pc and genetic interaction with Scm and esc.	*polyhomeotic proximal*
FBgn0004865	It is involved in instar larval or pupal development	*Ecdysone-induced protein 78C*
FBgn0004893	The gene is involved in lateral inhibition during wing disc development.	*brother of odd with entrails limited*
FBgn0004907	The gene is involved in Ras protein signal transduction	*14-3-3zeta*
FBgn0004913	The gene is involved in neurogenesis and lateral inhibition	*Germ line transcription factor 1*
FBgn0005533	The gene product is involved in translation.	*Ribosomal protein S17*
FBgn0005558	The gene regulates insulin-like growth factor	*eyeless*

	receptor signaling pathway	
FBgn0005564	The gene product is involved in synaptic transmission	*Shaker cognate l*
FBgn0005612	The gene is involved in dendrite morphogenesis (inferred from genetic interaction with Mical)	*Sox box protein 14*
FBgn0005677	The gene negatively regulates expression of other genes.	*dachshund*
FBgn0005771	It is involved in regulation of transcription from RNA polymerase II promoter, cell proliferation and wing disc development.	*no ocelli*
FBgn0010395	The gene is involved in wing morphogenesis and negative regulation of synaptic growth at neuromuscular junction.	*Integrin betanu subunit*
FBgn0010583	The gene is involved in insulin receptor signaling pathway.	*dreadlocks*
FBgn0010768	The gene is involved in and muscle development.	*squeeze*
FBgn0011584	SRP-dependent cotranslational protein targeting to membrane	*Translocation protein 1*

FBgn0011591	It is responsible for dorsal/ ventral pattern formation of wing disc, margin morphogeneis of the wing and is involved in Notch signalling pathway.	*fringe*
FBgn0011592	The gene product is responsible for axon guidance and dendrite morphogenesis.	*frazzled*
FBgn0011742	The gene regulates actin polymerization and depolymerization.	*Actin-related protein 2*
FBgn0011764	The gene negatively regulates transcription for RNA pol II promoter.	*Dorsal switch protein 1*
FBgn0011817	The gene product is involved in vein specification, and is a negative regulator of Wg signalling pathway.	*nemo*
FBgn0013469	The gene product is a negative regulator of ras signalling pathway.	*klumpfuss*
FBgn0013733	The gene product is responsible for apposition of dorsal and ventral imaginal disc-derived wing surfaces	*short stop*
FBgn0013983	The gene is involved in innate immune response.	*immune deficiency*
FBgn0014037	The gene product is involved in dorsal/ventral pattern formation of wing disc.	*Su(Tpl)*

FBgn0014163	The gene product is involved in axonogenesis.	*failed axon connections*
FBgn0015371	The gene regulates peripheral nervous system development.	*charlatan*
FBgn0015513	The gene is involved in cell morphogenesis and dorsal closure.	*myoblast city*
FBgn0015903	The gene isinvolved in negative regulation of JAK-STAT cascade.	*apontic*
FBgn0015919	The gene is involved in vein specification in wing imaginal discs.	*caupolican*
FBgn0016694	The gene product is involved inregulation of mitosis.	*PAR-domain protein 1*
FBgn0016930	protein phosphorylation | non-traceable author statement	*smell impaired 35A*
FBgn0016977	The gene is responsible for wing vein morphogenesis , maintenance of imaginal disc-derived wing hair orientation and wing discpattern	*split ends*

	formation. It shows interaction with EGFR signalling pathway and Wg signalling pathway.	
FBgn0019686	The gene acts at checkpoints of cell cycle.	*loki*
FBgn0020257	The gene product is responsible for ubiquitin-dependent protein catabolic process.	*partner of paired*
FBgn0020307	It is responsible for wing morphogenesis.	*defective proventriculus*
FBgn0020493	The gene is responsible for negative regulation of transforming growth factor beta receptor signaling pathway , anterior/posterior pattern specification of the wing disc.imaginal disc .	*Daughters against dpp*
FBgn0021768	nuclear migration	*nudC*
FBgn0022768	The gene product is responsible for protein dephosphorylation.	*Protein phosphatase 2C*
FBgn0023172	The gene is involved in cell elongation involved during wing morphogenesis.	*Rho guanine nucleotide exchange factor 2*

FBgn0024320	The gene is involved in regulation of cholesterol transport.	*Niemann-Pick type C-1a*
FBgn0024728		*Slip1*
FBgn0026077	chitin metabolic process | inferred from electronic annotation with InterPro:IPR002557	*Gasp*
FBgn0026179	The gene is involved in morphogeneis of wing imaginal disc.	*schizo*
FBgn0026181	The gene is involved in establishment of imaginal disc-derived wing hair orientation, regulation of cell shape and dorsal closure.	*Rho-kinase*
FBgn0026239	The gene product is involved in axon guidance.	*GUK-holder*
FBgn0026259	The gene product is involved in initiation of translation.	*eIF5B*
FBgn0026262	The gene is involved in positive regulation of transcription from RNA polymerase II promoter.	*bip2*
FBgn0026479	The gene product is involved in cytokinesis.	*Dynamin related protein 1*
FBgn0027339	The gene is involved in regulation of chromatin	*jim*

	silencing.	
FBgn0027865		*Tetraspanin 96F*
FBgn0027932	The gene regulates the establishment of planar polarity .	*A kinase anchor protein 200*
FBgn0028387	The gene is responsible for silencing other genes, interacts with Pc.	*chameau*
FBgn0028420	The gene is involved inmetamorphosis.	*Kruppel homolog 1*
FBgn0028494	-	-
FBgn0029003	The gene is involved in cell fate commitment.	*mab-21*
FBgn0029082	The gene is responsible for positive regulation of Notch signalling pathway.	*hibris*
FBgn0029114	Toll signaling pathway | inferred from electronic annotation with InterPro:IPR027202	*Tollo*
FBgn0029662	Electronic annotation ofthegene show that it is involved in cell redox reactions.	-

FBgn0029791	The gene is a negative regulator of EGFR pathway.	-
FBgn0029824	-	-
FBgn0029915	-	-
FBgn0030065	-	-
FBgn0030318	The gene product is responsible for proteolysis.	*rhomboid-4*
FBgn0030505	The gene is a negative regulator of EGFR pathway and is involved in wing morphogenesis.	*NFAT homolog*
FBgn0030719	The gene product is involved in neurogenesis.	*eIF5*
FBgn0030745	Based on electron annotation, the gene is predicted to be involved in monovalent inorganic cation transport.	-
FBgn0031037		-
FBgn0031057	The gene regulates protein metabolic process.	*Ubiquilin*
FBgn0031456	The gene regulates mRNA splicing via spliceosome (inferred from physical interactionwith B52, Rbp1, SC35, SF2 and x16).	*Transportin-Serine/Arginine rich*
FBgn0031488	RNA interference | inferred from mutant phenotype	-

FBgn0031668	The gene regulates neuron projection morphogenesis.	-
FBgn0031696	It acts a mitotic spindle assembly checkpoint protein.	*Bub1 homologue*
FBgn0031829	The gene is involved in microtubule-based movement.	*osm-6*
FBgn0031902	The gene is involved in Wnt receptor signaling pathway	*Wnt oncogene analog 6*
FBgn0032120	The gene product is involved in intracellular protein transport and golgi organization.	-
FBgn0032609	-	-
FBgn0032629	The gene is responsible for heterophilic cell-cell adhesion.	*beat-IIIc*
FBgn0032680	Based on sequence similarity it is predicted to play a role in transport of proteins to nucleus.	*Nuclear transport factor-2-related*
FBgn0032681	-	-
FBgn0033160	The gene is involved in neurogenesis. (inferred from mutant phenotype)	-

FBgn0033317		-
FBgn0033365	Based on electronic annotation, the gene is predicted to be associated with proteolysis.	-
FBgn0034483	-	-
FBgn0034590	-	*Magi*
FBgn0034657	-	*Lamin B receptor*
FBgn0034802		-
FBgn0034837	Based on structural similarity , the gene is predicted to be involved in translation.	*Ribosomal protein L22-like*
FBgn0034885	Based on sequence and structural similarity the gene is predicted to be assoiciated with transmembrane transport.	-
FBgn0034915	Based on sequence similarity, The gene is predicted to be involved in initiation of translation.	*eIF6*
FBgn0035101	The gene is involved in maintenance of cell polarity and cell-cell adhesion.	*p130CAS*
FBgn0035150	The gene product is involved in translesion	*Rev1*

	synthesis.	
FBgn0035411	The gene positively regulates cell size.	*Girdin*
FBgn0035529	-	-
FBgn0035696	Based on sequence similarity this gene is predicted otbe involved in anion transport.	*Bestrophin 2*
FBgn0036043		-
FBgn0036222	Based on electronic annotation the gene is predicited to be involved in tricarboxylic acid cycle .	-
FBgn0036302		*sosondowah*
FBgn0036446	Based on electronic annotation the gene is predicted to be involved in carbohydrate metabolic process.	-
FBgn0036490	The gene is responsible for response to DNA damage stimulus.	-
FBgn0036773	-	-
FBgn0036959	-	-
FBgn0037120	The gene is involved in regulating cell cycle.	-
FBgn0037153	The gene product is involved in oxidation-reduction process.	*olf413*

FBgn0037305	Based on from electronic annotation with InterPro:IPR010041 the gene is predicted to be involved in L-methionine salvage from methylthioadenosine.	-
FBgn0037341	The gene product is responsible for dephosphorylation of other proteins.	-
FBgn0037364	The gene product is involved in morphogenesis of a polarized epithelium..	*Rab23*
FBgn0037498	Based on sequence similarity with other proteins, the gene is predicted to be involved in protein folding.	-
FBgn0037705	The gene is involved in learning or memory.	*murashka*
FBgn0037718	-	*P58IPK*
FBgn0037973	Based on electronic annotation with other proteins, the gene is predicted to be involved in oxidation-reduction process.	-
FBgn0038167	The gene is involved in the process of dorsal closure and regulates JNK cascade.	*lkb1*
FBgn0038168	The gene is involved in neurogenesis.	*oocyte*

		maintenance defects
FBgn0038341	Based on electronic annotation with other proteins, the gene product is believed to be involved in proteolysis.	-
FBgn0038826	The gene product is involved in dorsal appendage formation.	*Syncrip*
FBgn0038881	phagocytosis, engulfment | inferred from mutant phenotype	-
FBgn0038983	The gene is involved in neurogenesis.	-
FBgn0039120	The gene product is responsible for import of SMAD protein into the nucleus.	*Nucleoporin 98-96kD*
FBgn0039131	-	-
FBgn0039266		-
FBgn0039633	-	-
FBgn0039728	-	-
FBgn0039902	Based on electronic annotation with other proteins the gene is predicted to be involved in transmemebrane transport.	-

FBgn0039907	The gene is responsible for wing margin morphogenesis and positively regulates Wnt signalling pathway.	*legless*
FBgn0040071	The gene product is responsible for dorsal/ventral pattern formation of the wing disc (inferred from genetic interaction with Bx, ap); chromatin-mediated maintenance of transcription (inferred from genetic interaction with brm, osa, pb, Pc, ph-p and trx. The gene product is also involved in lateral inhibition.	*taranis*
FBgn0040388	The gene product is involved in smoothened signalling pathway and also plays a role in wing disc pattern formation.	*brother of ihog*
FBgn0040752	The gene is involved in prosynaptic dendritic assembly.	*Prosap*
FBgn0041092	The gene product is involved in axon extension .	*taiman*
FBgn0041094	The gene acts as a negative regulator of growth (inferred from genetic interaction with chrb).	*scylla*
FBgn0041203	The gene is involved in establishment of imaginal disc-derived wing hair orientation.	*LIM-kinase1*
FBgn0041604	The gene is involved in Wnt receptor signaling pathway and decapentaplegic signaling pathway and	*dally-like*

	smoothened signaling pathway. The gene is also involved in wing disc dorsal/ventral pattern formation.	
FBgn0043903	The gene is involved in JAK-STAT cascade.	*domeless*
FBgn0050069	-	-
FBgn0051612	The gene product is involved in the process of lateral inhibition.	-
FBgn0052062	The gene is involved in wing vein specification.	*Ataxin-2 binding protein 1*
FBgn0052082	Based on electronic annotation, the gene is thought to be involved in filopodium assembly.	-
FBgn0052138	The gene product is involvedin neuron projection morphogenesis.	-
FBgn0052638	-	-
FBgn0053113	The gene is responsible for olfactory behaviour.	*Rtnl1*
FBgn0053474	Based on electronic annotation, thegene is predicted to be involved in peroxisome fission.	-
FBgn0066365	The gene is involved in cell-matrix adhesion.	*dusky-like*

FBgn0085424	The gene is involved inwing disc development and imparts identity to the growing wing blade.	*nubbin*
FBgn0085432	The gene negatively regulates Wg signllaing pathway.	*pangolin*
FBgn0085443	The gene is responsible for border follicle cell migration. (inferred from genetic interaction with Cbl, Egfr and Pvr.	*sprint*
FBgn0086613	The gene product regulates gene expression of other genes.	*Ino80*
FBgn0086655	The gene is responsible for wing vein specification.	*jing*
FBgn0086899	The gene is responsible for regulating cell cycle.	*Tousled-like kinase*
FBgn0086911	The gene is responsible for neuromuscular junction development.	*rugose*
FBgn0243512	The gene product is involved in eversion of imaginal disc and their fusion. The gene is also involved in JNK cascade.	*puckered*
FBgn0250839	-	-
FBgn0259211	The gene is responsible for regulation of cell shape and	*grainy head*

	regulation of transcription from RNA polymerase II promoter.	
FBgn0259220		
FBgn0259240	Based on structural similarity to other proteins the gene is predicted to be involved in cell adhesion.	*Tenascin accessory*
FBgn0259244		-
FBgn0259878	The gene is a negative regulator of Dpp signalling pathway.	*Follistatin*
FBgn0259984	The gene is involved in processing of Notch receptor.	*kuzbanian*
FBgn0260003	The gene product regulates wing vein morphogenesis.	*Dystrophin*
FBgn0260634	The gene product regulates meiotic G2/MI transition.	*eukaryotic translation initiation factor 4G2*
FBgn0260635	The gene is a positive regulator of canonical Wnt receptor signalling pathway.	*thread*
FBgn0260642	The gene is responsible for anterior/posterior axis specification.	*Antennapedia*

FBgn0260653	The gene is involved in open tracheal system development	*serpentine*
FBgn0261383	Based on structural similarity with other proteins, the gene is predicted to be involved in snRNA processing.	*Integrator 6*
FBgn0261439	The gene is a component of tricarboxylic acid cycle.	*Succinate dehydrogenase A*
FBgn0261570	The gene is involved in wing disc morphogenesis.	-
FBgn0261648	The gene is involved in wing pattering along the A-P axis, wing vein morphogenesis and notum fate specification.	*spalt major*
FBgn0261823	The gene is involved in sex comb development and chromatin silencing.	*Additional sex combs*
FBgn0261873	The gene is involved in establishment or maintenance of epithelial cell apical/basal polarity and lateral inhibition.	*stardust*
FBgn0262127	The gene regulates hippo signaling cascade (inferred from genetic interaction with ex, Mer, sav, hpo).	*kibra ortholog*
FBgn0262139	The gene is involved in lateral inhibition.	*trachealess*

FBgn0262614	The gene is involved in wing disc development and dorsal closure. The gene is also involved in JNK cascade.	*polychaetoid*
FBgn0262656	The gene is a positive regulator of growth and is involved in cell proliferation. It negatively regulates Notch signalling pathway.	*diminutive*
FBgn0262735	The gene is involved in nervous system development.	*IGF-II mRNA-binding protein*
FBgn0263097	The gene is responsible for wing disc morphogenesis.	*Glucose transporter 4 enhancer factor*
FBgn0004009	The gene is responsible for D-V patterning of the wing disc	*wingless*

Table 4: Biolgical processes associated with various common target genes of Ubx in hind wing of *Apis* and haltere of *Drosophila.*

Flybase Gene Id	**Associated Biological processes**	**Common name of the gene**
FBgn0000242		
FBgn0000308	The gene product is involved in cytokinesis, actomyosin contractile ring assembly and dorsal closure.	*chickadee*
FBgn0000479	The gene regulates courtship behavior and learning.	*dunce*
FBgn0000547	The gene product regulates lateral inhibition, dorsal appendage formation and regulates cell shape. Cell cell adhesion and wing morphogeneis are also regulated by the gene. The gene is a negative regulator of EGFR pathway. It negatively regulates hippo signaling pathway.	*echinoid*
FBgn0000575	The gene regulates cell proliferation, wing vein morphogenesis and lateral inhibition. | The gene also interacts with EGFR pathway.	*extra macrochaetae*
FBgn0000625	The gene is involved in eye and notum development.	*eyegone*
FBgn0000658	The gene is responsible for wing vein	*four-jointed*

	specification, establishment of wing hair orientation. The gene also interacts with Wnt signaling pathway.	
FBgn0001235	The gene product is responsible for proximal/distal pattern formation,wing and haltere morphogeneis.	*homothorax*
FBgn0001257	The gene is involved in cell adhesion and is a negative regulator or insulin receptor signaling pathway.	*Ecdysone-inducible gene L2*
FBgn0001323	The gene product is known to be involved in regulation of mitosis.	*knirps-like*
FBgn0002543	The gene product plays a role in axon guidance	*leak*
FBgn0002732	The gene is involved in notch signaling pathway and helps in cell fate specification.	*Enhancer of split malpha, Bearded family member*
FBgn0002733	The gene is involved in notch signaling pathway.	*Enhancer of split mbeta, helix-loop-helix*
FBgn0002735	The gene is involved in notch signaling pathway and helps in patterning the dorsal-	*Enhancer of split mgamma, helix-*

	ventral surface of the wing imaginal disc.	*loop-helix*
FBgn0003162	The gene is responsible for regulation of epithelial cell migration.	*Punch*
FBgn0003174	The gene is involved in lateral inhibition	*pawn*
FBgn0003310	The gene is involved in epidermal growth factor receptor signaling pathway. It helps in morphogeneisis of wing vein.	*Star*
FBgn0003371	The gene negatively regulates smoothened and wnt signaling pathway. It also regulates notch signaling pathway. It helps in subdivision of wing and notum fields and regulates epithelial cell planar polarity.	*shaggy*
FBgn0003396	The gene is responsible for cell proliferation, anterior- posterior pattern formation of the wing disc and helps in wing vein morphogenesis. It regulates TGF- beta signaling pathway.	*schnurri*
FBgn0003463	The gene is involved in wing vein morphogenesis	*short gastrulation*

	The gene regulates BMP and torso signaling pathway.	
FBgn0003525	The gene is involved in G2/M transition of mitotic cell cycle	*string*
FBgn0003862	The gene product is responsible for chromatin-mediated maintenance of transcription.	*trithorax*
FBgn0003870	The gene product is responsible for regulation of cell shape and wing morphogenesis.	*tramtrack*
FBgn0003997	The gene is involved in the process of apoptosis.	*Wrinkled*
FBgn0004101	The gene is responsible for apposition of dorsal and ventral wing surfaces.	*blistered*
FBgn0004449	The gene regulates cell shape and in involved in cell adhesion.	*Tenascin major*
FBgn0004509	The gene product is responsible for proteolysis.	*Furin 1*
FBgn0004603		

FBgn0004644	The gene is involved in anterior/posterior pattern formation in the wing disc. It also helps in specification of wing veins in the disc.	*hedgehog*
FBgn0004656	The gene is involved in wing morphogenesis.	*female sterile (1) homeotic*
FBgn0004854	The gene is involved in morphogenesis of chaeta and lateral inhibition.	*BarH2*
FBgn0004914	The gene product is involved in ectoderm and mesoderm development.	*Hepatocyte nuclear factor 4*
FBgn0005630	The gene is involved in the process of axon guidance.	*longitudinals lacking*
FBgn0005634		
FBgn0005672	The gene is involved in EGFR pathway. It regulates mitotic cell cycle and apopotosis.	*spitz*
FBgn0010113	The gene is involved in cell differentiation and affects wing morphogeneisis.	*headcase*
FBgn0010300	The gene is a negative regulator of cell proliferation.	*brain tumor*

FBgn0010313	The gene is involved in chromosome condensation.	*corto*
FBgn0010379	The gene is a positive regulator of cell and organ growth. It is involved in insulin receptor signaling pathway.	*Akt1*
FBgn0010438	The gene is responsible for maintainence of mitochondrial genome.	*mitochondrial single stranded DNA-binding protein*
FBgn0010548	The gene is involved in oxidation reduction process.	*Aldehyde dehydrogenase type III*
FBgn0010575	The gene regulates imaginal disc growth and is involved in wing morphogenesis . It is a negative regulator of smoothened signaling pathway.	*scribbler*
FBgn0010762	The gene is responsible for specification of cardiac cell fate.	*simjang*
FBgn0011224	The gene is involved in notch signaling pathway.	*hephaestus*

	It plays a role in wing margin and wing vein morphogenesis.	
FBgn0011236	The gene regulates JNK cascade.	*ken and barbie*
FBgn0011335		*lethal (3) j2D3*
FBgn0011661	The gene is responsible for establishment and maintenance of epithelial cell apical/ basal polarity. It is also involved in anterior/ posterior pattern specification.	*Moesin*
FBgn0011725	The gene is involved in shortening of poly A tail of mRNA.	*twin*
FBgn0011766	The gene is a positive regulator of cell proliferation The gene is also involved in wing morphogenesis.	*E2F transcription factor*
FBgn0011837	The gene product negatively regulates expression of other genes.	*Tis11 homolog*
FBgn0013263	The gene is involved in chromatin	*Trithorax-like*

	organization and wing morphogeneis.	
FBgn0014019	The gene prduct is responsible for phototransduction.	*Rhodopsin 5*
FBgn0015229	The gene is responsible for heterophilic cell-cell adhesion.	*gliolectin*
FBgn0015541	The gene affects mechanosensory behavior of flies	*slamdance*
FBgn0016059	The gene is responsible for axon guidance.	*Sema-1b*
FBgn0016715	Based on electronic annotation this gene is thought ot be involved in metabolic process.	*Rhythmically expressed gene 2*
FBgn0016754	-	*six-banded*
FBgn0016797	The gene is involved in wnt signaling pathway. The gene is involved in wing margin morphogenesis.	*frizzled 2*
FBgn0019968	The gene is involved in establishment of spindle orientation and regulation of synapse structure and activity	*Kinesin heavy chain 73*
FBgn0020245		

FBgn0020280	The gene is involved in embryonic development	*labial associated factor*
FBgn0022764	The gene product is involved in regulation of mitotic cell cycle.	*Sin3A*
FBgn0023001	The gene in involved in insulin receptor signaling pathway	*melted*
FBgn0023215	The gene is a negative regulator of cell growth and also regulates cell cycle.	*Mnt*
FBgn0023526	regulation of cell cycle | inferred from direct assay	-
FBgn0024250	The gene regulates BMPsignalling pathway and is responsible for wing disc morphogenesis.	*brinker*
FBgn0024289	Based on electronic annotation this gene is predicted to play a role in oxidation-reduction process.	*Sorbitol dehydrogenase 1*
FBgn0024321	Based on sequence similarity this gene is predicted to regulate transcription.	*NK7.1*
FBgn0025390	Based on electronic annotation this gene is	*Mucin related 2B*

	predicted to play a role in chitin metabolic process	
FBgn0025574	The gene is involved Toll signaling pathway and immune response.	*Pellino*
FBgn0026160	The gene is involved in chromatin-mediated maintenance of transcription.	*tonalli*
FBgn0026263		*bip1*
FBgn0027106	The gene is involved in wing disc morphogenesis.	*Innexin 7*
FBgn0027108	The gene has a role to play in olfactory behavior and intracellular transport.	*Innexin 2*
FBgn0027343	The gene is involved in Wg signaling pathway. The gene is responsible for establishment or maintenance of cell polarity	*frizzled 3*
FBgn0027529	-	-
FBgn0027546	Based on electronic annotation this gene is predicted to play a role in cell fate commitment	-

FBgn0027621	The gene is involved in wing morphogeneis	*6-phosphofructo-2-kinase*
FBgn0028421	The gene is involved in establishment or maintenance of microtubule cytoskeleton polarity	*Kinesin associated protein 3*
FBgn0028686	The gene is involved in response to DNA damage stimulus	*Regulatory particle triple-A ATPase 3*
FBgn0029504	The gene is a regulator of cell division.	*Checkpoint suppressor homologue*
FBgn0029831	The gene is involved in wing morphogeneis.	-
FBgn0030049	The gene is involved in ncRNA polyadenylation involved in polyadenylation-dependent ncRNA catabolic process	-
FBgn0030844	The gene is involved in male courtship behavior and inter male aggressive behavior.	*pickpocket 23*
FBgn0031429		
FBgn0031590	-	-
FBgn0031850	The gene is involved in cell adhesion	*Thrombospondin*

	mediated by integrin.	
FBgn0032694	The gene is a negative regulator of Ras signaling pathway.	*Misexpression suppressor of ras 3*
FBgn0032724		-
FBgn0033521		-
FBgn0034082		-
FBgn0034091	Based on electronic annotation this gene is predicted to play a role in protein folding.	*mrj*
FBgn0034321	-	-
FBgn0034408	The gene is responsible for regulation of tube length of open tracheal system	*serrano*
FBgn0034481	-	-
FBgn0034500	Based on electronic annotation this gene is predicted to play a role in metabolic process.	*Carbonyl reductase*
FBgn0034501	-	-

FBgn0034718	The gene is responsible for trachea development and synaptic target recognition	*windpipe*
FBgn0034909	Based on electronic annotation this gene is predicted to play a role in transmembrane transport.	-
FBgn0035142	The gene is involved in positive regulation of Notch signaling pathway It is also a positive regulator of Wnt and hippo signaling pathway. It is involved inwing vein morphogeneisis.	*homeodomain interacting protein kinase*
FBgn0035161	-	-
FBgn0035162	The gene is involved in organization of mitotic spindle.	-
FBgn0035232	-	-
FBgn0035233	The gene product is responsible for primary spermatocyte growth and spermatocyte division	*Peroxin 10*
FBgn0035426	The gene is involved in neuron projection	-

	morphogenesis	
FBgn0035445	Based on electronic annotation this gene is predicted to play a role in metabolic process.	-
FBgn0035533	The gene plays a role in wing hair organization	*Cip4*
FBgn0035954	The gene plays a role in cardiocyte differentiation and anterior malpighian tubule development.	*Dorsocross3*
FBgn0036112	-	-
FBgn0036154	Based on structural similarity to other proteins, this gene is predicted to play a role in proteolysis.	-
FBgn0036156	-	-
FBgn0036373	-	-
FBgn0036577	The gene product is involved in neurogenesis.	-
FBgn0036782	-	-
FBgn0036783	Based on structural similarity to other proteins, this gene is predicted to play a role in sensory perception of chemical stimulus	*Chemosensory protein A 75a*

FBgn0036900	The gene playas a role in lateral inhibition.	-
FBgn0037350	-	-
FBgn0037645	-	-
FBgn0037659	The gene plays a role in segment specification.	*Lysine (K)-specific demethylase 2*
FBgn0037720	-	-
FBgn0037856	-	-
FBgn0038188	Based on electronic annotation, this gene is predicted to play a role in protein methylation	*Arginine methyltransferase 9*
FBgn0038919		-
FBgn0039155	-	*Kallmann syndrome 1 ortholog*
FBgn0039350	Based on structural similarity to other proteins, this gene is predicted to play a role in regulating gene expression	*jing interacting gene regulatory 1*
FBgn0039669	Based on structural similarity to other proteins, this gene is predicted to play a role	-

	in mitochondrial electron transport, NADH to ubiquinone	
FBgn0039804	-	-
FBgn0039864	-	-
FBgn0039958	-	-
FBgn0040066	The gene product acts as mitotic G2 DNA damage checkpoint protein.It is also involved in chromatin remodeling.	*will die slowly*
FBgn0042094	Based on structural similarity to other proteins, this gene is predicted to play a role in ADP biosynthetic process	*Adenylate kinase-3*
FBgn0044823	Based on structural similarity to other proteins, this gene is predicted to play a role in regulation of Rho protein signal transduction	*Spec2*
FBgn0046763	The gene is a negative regulator of Wnt signaling pathway	-
FBgn0050447		-
FBgn0051163	The gene is involved in sensory perception of smell.	*Shal K[+] channel interacting protein*

FBgn0051721	The gene is involved in axiogenesis and axon guidance.	*Trim9*
FBgn0052264	-	-
FBgn0052529	The gene plays a role in chromatin silencing and remodeling. It is also responsible for wing morphogenesis.	*Histone gene-specific Epigenetic Repressor in late S phase*
FBgn0052676	The gene plays a role in lateral inhibition.	-
FBgn0053196	The gene is involved in apposition of dorsal and ventral surfaces of the wing disc. It is also involved in lateral inhitbion.	*dumpy*
FBgn0053648		-
FBgn0053796		-
FBgn0054006		-
FBgn0054046		-
FBgn0054050		-
FBgn0063261	The gene product is involved in lateral	-

	inhibiton.	
FBgn0083919		
FBgn0083940		-
FBgn0083978	-	-
FBgn0085201	-	-
FBgn0085318	-	-
FBgn0085397	The gene regulates apoptosis and wing morphogenesis.	*Fish-lips*
FBgn0085403		
FBgn0085450	The gene is negative regulator of TGF beta signaling pathway.	*Sno oncogene*
FBgn0085489	The gene is responsible for wing morphogenesis.	-
FBgn0086902	The gene is responsible for wing morphogenesis.	*kismet*
FBgn0250823	The gene is involved in Wnt signaling pathway.	*gilgamesh*

FBgn0259789	The gene regulates mitotic cell cycle and cellularization.	*vielfaltig*
FBgn0259986	The gene plays a role in pattern formation in the wing disc and cell proliferation.	*nab*
FBgn0261284	The gene plays a role in assembly of septate junctions.	*boudin*
FBgn0261885	It regulates EGFR pathway and interacts with Wg signaling pathway. The gene plays a role in wing margin and wing vein morphogenesis	*osa*
FBgn0262736	The gene is involved in wing morphogenesis.	*Vacuolar H[+] ATPase 16kD subunit 1*

Table 5: Ubx target genes specific to diptera and their biological role in the fly

Motif	1	2	3	4	5	6
Ubx (motif from TRANSFAC)	0.333	0.165	0.4955	0.392	0.477	1.2168
Ubx (motif from Noyes *et. al.*)	1.933	0.769	0.3978	1.122	1.148	1.023
Ubx (motif from Mann *et. al.*)	16.9	7.783	0.4605	7.936	7.964	1.003
Pend	0.043	0.018	0.4186	0.032	0.049	1.53
TCF	0.687	0.421	0.6128	0.464	0.477	1.028
MAD	0.591	0.531	0.8985	0.496	0.453	0.913
dElF1	0.064	0.055	0.8594	0.099	0.089	0.8989
Snail	0.354	0.311	0.8785	0.442	0.404	0.914
GAGA	5.798	6.172	1.0645	0.816	0.97	1.1887
MAZ	0.29	0.641	2.2103	0.054	0.097	1.796
E2f	0.472	0.696	1.4746	0.37	0.259	0.7
Adf	0.107	0.201	1.8785	0.108	0.097	0.898
Myc	0.011	0.055	5	0.014	0.008	0.5714

Table 6: Table showing the preference of motif utilisation by hymenopterans and dipterans.

Column 1: Frequency of motifs in honey bee sequences corresponding to genes specific to *Apis mellifera.*

Column 2: Frequency of motifs in honey bee sequences corresponding to genes common to *Apis mellifera* & *Drosophila melanogaster.*

Column 3: Over representation of motifs in common sequences when compared to specific sequences of honey bee.

Column 4: Frequency of motifs in fruit fly sequences corresponding to genes specific to *Drosophila melanogaster.*

Column 5: Frequency of motifs in honey bee sequences corresponding to genes common to *Apis mellifera* & *Drosophila melanogaster.*

Column 6: Over representation of motifs in common sequences when compared to specific sequences of fruit fly.

Appendices

Appendix

Chapter 2

2.12: Generation of polyclonal antibodies against recombinant Ubx protein of *Apis mellifera*

Step3: Amplification and ligation of N- terminal region of Ubx

Forward Primer: 5'GGG AAT TCC ATA TGT ATT TTG AGC AGA CTG CG 3'

Reverse Primer: 5'CGC GGA TCC GTT GTT GCC AGG ACT CGA 3'

Step 4: Expression and purification of recombinant protein

Composition of Terrific broth

Tryptone	:	12g
Yeast Extract	:	24g
Glycerol	:	4mL

The components were dissolved in 900mL and autoclaved for 15min at 15 psi on liquid cycle.

When the solution cooled to 60°C , 100 mL of a sterile solution of 0.17M KH_2PO_4, 0.72 M K_2HPO_4 was added to broth (this solution was made by dissolving 2.31 grams of KH_2PO_4 and 12.54 grams of K_2HPO_4 in 90 mL of deionized water. After the salts dissolved , the volume of the solution was adjusted to 100 mL with deionized water and sterilized by autoclaving for 20 minutes at 15 psi on liquid cycle).

Composition of denaturing lysis buffer

NaH_2PO_4	:	100mM
Tris	:	10mM
Urea	:	8M
pH	:	8

Step5: Generation of polyclonal antibodies against the recombinant protein

Composition of running buffer

Tris base	:	3.03 g
Glycine	:	14.4 g
SDS	:	1.0 g
Total volume	:	1000mL

Composition of 2X loading dye

Glycerol	:	2.5mL
10%SDS	:	4.0 mL
Bromophenol blue	:	0.1%
Beta- Mercaptoethanol:		0.8 mL
Milli Q water	:	0.7 mL
Total volume	:	10.0mL

Composition of staining solution

Coomassie brilliant blue (R-250)	:	0.3 g
Methanol	:	80mL
Acetic acid	:	20mL
Water	:	100mL

After adding CBB to methanol, the solution was stirred for one hour. Acetic acid & water were then added to the solution which was filtered and stored in dark bottles.

Composition of destaining solution

Acetic acid	:	100mL
Methanol	:	300mL
Water	:	600mL

Step 6: Testing the specificity of antibody

Composition of transfer buffer

Tris	:	18.2 g
Glycine	:	86.5g
Methanol	:	600 mL
SDS	:	0.1%
Total volume	:	4 lites with Milli Q

Composition of TBST

Tris HCl (50mM)	:	6.05 g

Sodium chloride (150mM) : 8.76 g

pH 7.5

Tween 20 : 1mL

diluted to 1000 mL

Composition of insect cell lysis buffer

Tris Cl. : 100 mM , pH 7.5

Sodium Chloride : 130 mM

Triton X- 100 : 1%

Sodium fluoride : 10mM

Sodium phosphate : 10mM ,pH 7.5

Sodium pyrophosphate: 10mM

The solution was filter sterilised.

2.13: Chromatin Immunoprecipitation

Swelling buffer

25mM Hepes, pH 7.9

1.5mM MgCl2

10mM KCl

0.1% NP40

1mM DTT

0.5mM PMSF

Protease inhibitors

Sonication buffer

50mM Hepes , pH 7.9

140mM NaCl

1mM EDTA

1% Triton X-100

0.1% sodium deoxycholate

1% SDS

0.5mM PMSF

Protease inhibitor cocktail

Dilution buffer

50mM Hepes , pH 7.9

140mM NaCl

1mM EDTA

1% Triton X-100

0.1% sodium deoxycholate

0.5mM PMSF

Protease inhibitor cocktail

Wash buffer A

50mM Hepes , pH 7.9

500mM NaCl

1mM EDTA

1% Triton X-100

0.1% sodium deoxycholate

1% SDS

0.5mM PMSF

Protease inhibitor cocktail

Wash buffer B

20mM Tris,ph 8

1mM EDTA

250mM LiCl

0.5 % NP40

0.5% sodium deoxycholate

0.5mM PMSF

Protease inhbotor cocktail

Elution buffer

50mM Tris,pH 8

1mM EDTA

1% SDS

50mM $NaHCO_3$

Appendix

Chapter 4

4.1: Validation of peaks (using ChIP-qPCR)

List of primers

RplP0_FP: 5’- CCCAGGTTAAAGGACTTCCCTCGGA- 3’

RplP0_RP: 5’- TCGGTTTGAATCGATCTTCCGCGA -3’

AmChIP46_FP: 5’- CGGCGTGGCAGCGTAGGTAT -3’

AmChIP46_RP: 5’- GTTGGCCGGTCGGCAAGGTT -3’

AmChIP47_FP: 5’- GGTGCGCGGTGTATCCTCGG-3’

AmChIP47_RP: 5’- CACGGCCCCGGGTGATTACG -3’

AmChIP48_FP: 5’- CGCCCGCTAGAGCGCATGTT -3’

AmChIP48_RP: 5’ – GCAGACGCCACCCTTGCGAT -3’

AmChIP26_FP: 5’- AGGTGGGGGCTGTGGCCAAT – 3’

AmChIP26_RP: 5’- TGCGAATTGGAAGGAGGCGGC- 3’

AmChIP27_FP: 5’- GGATCCCCTGGCCCCTCTGG -3’

AmChIP27_RP: 5’- TGGCGAGCTAGCTGGTGGGT -3’

AmChIP33_FP: 5’- CGGCGTGGCAGCGTAGGTAT -3’

AmChIP33_RP: 5’- TTGGCCGGTCGGCAAGGTTT- 3’

In all the subsequent PCRs, the primers were used at a dilution of 2 pm/ µl. The reaction was carried out on Eppendorf real time machine. In all the reactions described below, amplification was carried out using input, anti Ubx and pre immune chipped DNA sample. No template control was always used on the same real time plate to detect false positives, if any. A melting curve was also included in the end to check the presence of primer dimers.

Finally enrichment was calcµlated against the input for both the anti Ubx chipped DNA samples and the pre immune chipped DNA samples and plotted.

Reaction mix for all the real time reactions for validation of chipped peaks (please note that the primers used are different in each case, depending on the peak to be validated and the template can be input, chipped or pre immune Iped DNA.)

MQ	:	2.0 µl
FP	:	1.0 µl
RP	:	1.0 µl
Template	:	1.0 µl
Mesa green mix	:	5.0 µl
Total volume	:	10.0 µl

General reaction conditions for amplification of various peaks (please note that the Tm of all the primers used was 60ºC)

95 ºC for 5 minutes

Followed by 40 cycles of

95 ºC for 15 seconds

60 ºC for 30 seconds

72 ºC for 30 seconds

Followed by a melt curve of

95 ºC for 15 seconds

60 ºC for 15 seconds

Following which temperature was raised to 95 ºC over a time window of 20 minutes.

4.3: Validation of few Ubx target genes in *Apis mellifera*

List of primers:

GB 13747RT_FP: 5'- CCCCGTTATGTCCCCGCACC-3'

GB 13747RT_RP: 5' – CGGTCATCCGTCCCCTGGGC-3'

Tm: 60 ºC

GB11410RT_FP: 5'- GCAGTCGAGAGCACCCCGGA-3'

GB11410RT_RP: 5'- CTTCGCTTGGGACGGCTGGG-3'

Tm: 60 ºC

GB17325RT_FP: 5'- TTCCTGGTGGTAAAGTACATGG-3'

GB17325RT_RP: 5'- ATACTGAATACGTCTCTTAGCTCG-3'

Tm: 55 ºC

GB10329RT_FP: 5’- AAATGAGTAAGGCTCATCCTC- 3’

GB10329RT_RP: 5’- CCACGTATTACCACCATTCC-3’

Tm: 55 ºC

GB15719RT_FP: 5’- CCCAAATACAGAACTTACTTGCCAG -3’

GB15719RT_RP: 5’- TTGTTGATTAGCCATTTGATGCCTC-3’

Tm: 55 ºC

Am_RpLP0RT_FP: 5’- TTAAGATGGGTAGGGAGGACAAGG-3’

Am_RpLP0RT_RP: 5’- TGTCTGCACCCACAATGAAACAC- 3’

Tm: 60 ºC

Real time PCR was carried out using the following mix and conditions. Different Tm was used for different primers which has been specified along with the primer sequence.

Reaction mix for amplification

MQ	:	2.0 µl
FP	:	1.0 µl
RP	:	1.0 µl
Template (cDNA/genomic DNA/water)	:	1.0 µl
Mesa green mix	:	5.0 µl
Total volume	:	10.0 µl

Reaction conditions for amplification

95 ºC for 5 minutes

Followed by 40 cycles of

95 ºC for 15 seconds

Tm (primer dependent) for 30 seconds

72 ºC for 30 seconds

Followed by a melt curve of

95 ºC for 15 seconds

60 ºC for 15 seconds

Following which temperature was raised to 95 ºC over a time window of 20 minutes.

Appendix

Chapter 6

6.11: Generation of fμll length *Gliolectin* clone

Primers Used:

FP: 5’- GCGGAAAGAACAAAGC- 3’

RP: 5’- TTGGACCGTCAAGTCTTGGT -3’

The following mix was used to get the amplicon of *Gliolectin*:

Milli Q	:	17.0 μl
10X Buffer		
(Invitrogen Pfx polymerase)	:	2.5 μl
Mg (50mM)	:	0.5 μl (2.5mM final conc)
FP (10pm/μl)	:	1 μl
RP(10pm/μl)	:	1 μl
dNTP	:	1μl
Template (cDNA)	:	1μl
DMSO	:	0.5μl
Pfx polymerase (Invitrogen)	:	0.5 μl
Total volume	:	25.0 μl

Program used for this amplification was:

95 ºC for 5 minutes followed by

30 cycles of

95 ºC for 30 seconds

55 ºC for 30 seconds

68 ºC for 1 minute

Followed by

Final amplification at 72 ºC for 5 minutes.

The amplicon was A – tailed after purification using the following mix:

Mg (50mM)	:	0.5µl (2.5mM final conc)
Taq	:	1.0µl
dATP	:	0.4µl
Buffer	:	1.0 µl
MQ	:	3.1µl

The mix was incubated at 70 ºC for 30 minutes.

This tailed mix (ratio of vector: insert being 1:3) was used for ligation using the following reaction mix

2X ligase buffer	:	5µl
pGEM- T easy vector	:	1µl
PCR product	:	1µl

T4 DNA ligase	:	1.0 µl
Mq	:	2.0 µl

It was incubated at 4 C overnight. The mix was used to transform DH 5 alpha cells which were selected on ampicllin plates (100 µg/ mL).

6.12: Preparation of *Gliolectin* probe

In order to generate template for in vitro transcription, the following program was used:

MQ water	:	17.5 µl
10 X pfu buffer	:	2.5 µl
$MgCl_2$	:	0.5 µl (final con 2.5mM)
M13 Forward primer (10pm/µl)	:	1.0 µl
M13 Reverse primer (10 pm/µl)	:	1.0 µl
Glec in pGEM T easy	:	1.0 µl
dNTP mix (10 mM each)	:	1.0 µl
Pfu Polymerase	:	0.5 µl

Expected amplicon size : 1225bp

Program used:

95 ºC for 5 minutes

Followed by 30 cycles of

95 ºC for 30 s

55 ºC for 30 s

72 ºC for 1 minute 30 sec

Followed by a final amplification at

72 ºC for 5 minutes

For preparation of RNA probe the following program was used:

DNA template (prepared as above)	:	3 µg
10 x Transcription buffer (Roche)	:	2 µl
10 X DIG- labelled dNTP mix (Roche)	:	2µl
RNAse inhibitor (Roche)	:	1µl
SP6 polymerase (Roche)	:	2µl

Milli Q to total volume 20 µl

It was incubated at 37 ºC for 3 hours in a thermal cycler – lid at 42 ºC.

The reaction was stopped by adding 2 µl of 200 mM EDTA.

6.13: In situ hybridisation in imaginal discs of *Drosophila melanogaster*

Composition of AP buffer

Tris Cl	:	100mM
NaCl	:	100 mM
Mg Cl_2	:	50mM

pH 9.5

The buffer was always prepared fresh

Composition of hybridisation buffer

5 X SSC pH 5

Deionised Formamide: 50 %

Tween 20 – 0.2 %

50 µg/ml salmon sperm DNA

50 µg/ml heparin

Composition of SSC 20 X

8.75g NaCl

4.41 g sodium citrate , pH 5.0 with HCl

6.14:Generation of Transgenics

6.14.1: Generation of UAS- *Gliolectin* construct

The following primer pair was used:

FP: CCG GAA TTC GCG GAAA GAA GAAA CAAA GC

RP: CCG CTC GAG TTG GAC CGT CAA GTC TTG GT

The expected amplicon size when using this primer was 995 bp

The mix used for amplification was

Milliq : 17.5 µl

10 X Pfx polymerase buffer : 2.5 µl

$MgCl_2$	:	0.5 µl (final conc 2.5 mM)
Forward Primer (10pm/µl)	:	1µl
Reverse Primer (10pm/µl)	:	1µl
dNTP mix (10mM each)	:	1µl
Template (glec in pGEM- T easy vector):		1µl
Pfx Polymerase	:	0.5 µl
Total volume	:	25 µl

Program Used for amplification was:

95 C for 5 minutes

Followed by 30 cycles of

95 ºC for 30 sec

55 ºC for 30 sec

68 ºC for 1 minute

Followed by final elongation of

72 ºC for 5 minutes

And the reaction mix was stored at 4 ºC

The amplified product was then electrophoresed on an agarose gel and the corresponding band was excised and DNA eluted .

The DNA was then restricted using ECoR1 and Xho1 and ligated in pUAST vector. (insert :vector ratio :: 3:1)

Mix used for restricting the amplicon

DNA	:	10 µl
10 x buffer (NEB ECoR1 buffer):		5µl
BSA	:	5µl
ECoR1	:	1µl
Xho1	:	1µl
MQ	:	28 µl

Total : 50 µl

Mix was incubated at 37 ºC for 3 hours. The restricted DNA was electrophoresed on gel – desired fragments eluted, mixed in the ratio mentioned above and were ligated at 4 ºC overnight and the ligated product was used for transformation.

Mix used for ligation:

Vector: 7.0 µl (200 ng)

Insert: 1.5 µl (66 ng)

10 x ligase buffer (promega): 1.6 µl

T4 DNA ligase: 1.0 µl

Mq to make volume to 16 µl

6.14.2: Construction of UAS RNAi *Gliolectin* construct

Construct1

In order to synthesize the construct 1 primers corresponding to the regions were generated and xho1 and Not 1 site was added to forward primer and reverse primer respectively.

FP: CCGCTCGAGTGTTGTGTCCGCCAATGGC

RP: ATAAGAATGCGGCCGCGATGTCATATC CGTAGCGTTC

Expected amplicon size 222 bp

The follwing mix was used for amplification:

MQ	:	17.5 µl
10 X Pfx buffer(invitrogen)	:	2.5µl
Mg	:	0.5µl (final conc 2.5 mM)
FP(10pm/µl)	:	1µl
RP (10pm/µl)	:	1µl
dNTP mix (10mM each)	:	1 µl
Template (glec in pGEM- T easy):		1µl
pfx polymerase	:	0.5 µl

total 25 µl

Program used for amplification

95 ºC for 5 minutes

Followed by 30 cycles of

95 ºC for 30 seconds

55 ºC for 30 seconds

68 ºC for 30 seconds

Followed by final amplification at 72 ºC for 5 min

Following amplification the PCR product was electrophoresed on an agarose gel and the band of desired size was eluted out.

Both, the excised PCR product and SympUAST vector were double digested using Not1 and Xho1 using the mix below

DNA : 5 µl

10 X buffer (NEB buffer 3) : 5 µl

BSA : 5 µl

Not1 : 1µl

Xho1 : 1µl

MQ : 33µl

Total: 50 µl

The mix was incubated at 37 ºC for 3 hours. After which it was electrophoresed on an agarose gel and the desired band was excised and eluted .

The insert and vector were mixed in a ratio of 3:1 and ligated overnight at 4 ºC

The following mix was used for ligation

Vector	:	200 ng
Insert	:	10 ng
10 X ligase buffer (Promega)	:	1.6µl
T4 DNA ligase	:	1.0 µl
Mq:	to make volume to 16µl	

The mix was used to transform bacteria which were selected on ampicillin plates- the plasmid was then sequenced to verify the insert.

Construct 2

In Order to prepare construct 2 primers were synthesized corresponding to the region mentioned and sites for ECoR1 was added to FP while that of Xho1 was added to RP.

The following primers were used for amplification:

FP: 5'-CCGGAATTCGACCTCAAGGACGATATCCAGCAC -3'

RP: 5'-CCGCTCGAGGCTTCTGATCTGCCAATTCGCTAG-3'

The follwing mix was used for amplification:

MQ	:	17.5 µl
10 X Pfx buffer (Invitrogen)	:	2.5µl
Mg	:	0.5µl (final conc 2.5 mM)
FP(10pm/µl)	:	1µl
RP (10pm/µl)	:	1µl

dNTP: 1 µl

template: 1µl

pfx polymerase : 0.5 µl

toal 25 µl

Program used for amplification

95 ºC for 5 minutes

Followed by 30 cycles of

95 ºC for 30 seconds

55 ºC for 30 seconds

68 ºC for 30 seconds

Followed by final amplification at 72 ºC for 5 min

Following amplification the PCR product was electrophoresed on an agarose gel and the band of desired size was eluted out.

Both, the excised PCR product and SympUAST vector were double digested using ECoR1 and Xho1 using the mix below

DNA : 5 µl

10 X buffer (NEB ECoR1 buffer) : 5 µl

BSA : 5 µl

ECoR1 : 1µl

Xho1 : 1µl

MQ : 33µl

Total : 50 µl

The mix was incubated at 37 ºC for 3 hours. After which it was electrophoresed on an agarose gel and the desired band was excised and eluted .

The insert and vector were mixed in a ratio of 3:1 and ligated overnite at 4 ºC

The following mix was used for ligation

Vector : 200 ng

Insert : 20 ng

10 X ligase buffer (promega): 1.6µl

T4 DNA ligase : 1.0 µl

MQ : to make volume to 16µl

The mix was used to transform bacteria which were selected on ampicillin plates- the plasmid was then sequenced to verify the insert

Composition of injection buffer (10X)

1mM sodium phosphate buffer pH 6.8

50mM KCl

Filter through 0.22 u filter- store in -20ºC

References

Abouheif, Ehab, and Gregory A Wray. Evolution of the Gene Network Underlying Wing Polyphenism in Ants. *Science* 297 (2008): 249–252.

Agrawal P, Habib F., Yelagandula R and Shashidhara LS. Genome level identification of targets of hox protein Ultrabithorax in Drosophila : Novel methods of target selection. *Scientific Reports* 205 (2011):

Akam, M. Hox Genes and the Evolution of Diverse Body Plans. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 349.1329 (1995): 313–9.

Akam, M. Hox Genes, Homeosis and the Evolution of Segment Identity: No Need for Hopeless Monsters. *The International journal of developmental biology* 42.3 (1998): 445–51.

Akam, M E, and a Martinez-Arias. The Distribution of Ultrabithorax Transcripts in Drosophila Embryos. *The EMBO journal* 4.7 (1985): 1689–700.

Alonso, Claudio R, and Adam S Wilkins. The Molecular Events That Underlie Developmental Evolution. *Nature Reviews Genetics* 6.September (2005): 709–715.

Angelini, David R, and Thomas C Kaufman. Insect Appendages and Comparative Ontogenetics. *Developmental biology* 286.1 (2005): 57–77.

Ansorge, Wilhelm J. Next-generation DNA Sequencing Techniques. *New biotechnology* 25.4 (2009): 195–203.

Appel, B, and S Sakonju. Cell-type-specific Mechanisms of Transcriptional Repression by the Homeotic Gene Products UBX and ABD-A in Drosophila Embryos. *The EMBO journal* 12.3 (1993): 1099–109.

Aspland, Simon E, and Robert A H White. Nucleocytoplasmic Localisation of Extradenticle Protein Is Spatially Regulated Throughout Development in Drosophila. *Development* 124 (1997): 741–747.

Averof, M, and N H Patel. Crustacean Appendage Evolution Associated with Changes in Hox Gene Expression. *Nature* 388.6643 (1997): 682–6.

Averof, Michalis. Arthropod Evolution: Same Hox Genes , Different Body Plans Surveys of Hox Genes in Various Arthropods and Related. *Current Biology* 7 (1997): 634–636.

Averof M. Arthropod Hox Genes: Insights on the Evolutionary Forces That Shape Gene Functions. *Current opinion in genetics & development* 12.4 (2002): 386–92.

Baeg, Gyeong-Hun et al. The Wingless Morphogen Gradient Is Established by the Cooperative Action of Frizzled and Heparan Sulfate Proteoglycan Receptors. *Developmental biology* 276.1 (2004): 89–100.

Bajpai, Ruchi, Kalpana Makhijani, et al. Drosophila Twins Regulates Armadillo Levels in Response to Wg/Wnt Signal. *Development* 131.5 (2004): 1007–16.

Bajpai, Ruchi, Nagraj Sambrani, et al. Identification of a Novel Target of D/V Signaling in Drosophila Wing Disc: Wg-independent Function of the Organizer. *Gene expression patterns : GEP* 5.1 (2004): 113–21.

Baraate, Sebastien, Christian Peeters, and Jean S Deutsch. Testing Homology with Morphology, Development and Gene Expression: Sex-specific Thoracic Appendages of the Ant Diacamma. *Evolution & Development* 8.5 (2006): 433 –445.

Beachy, P a et al. Cooperative Binding of an Ultrabithorax Homeodomain Protein to Nearby and Distant DNA Sites. *Molecular and cellular biology* 13.11 (1993): 6941–56.

Bennett, R L, S J Brown, and R E Denell. Molecular and Genetic Analysis of the Tribolium Ultrabithorax Ortholog, Ultrathorax. *Development genes and evolution* 209.10 (1999): 608–19.

Berger, Christian et al. A Critical Role for Cyclin E in Cell Fate Determination in the Central Nervous System of Drosophila Melanogaster. *Nature cell biology* 7.1 (2005): 56–62.

Berger, Christian et al. Cyclin E Acts Under the Control of Hox-Genes as a Cell Fate Determinant in the Developing Central Nervous System. *Cell cycle* 4.March (2005): 422–425.

Beye, M, and M Hasselmann. Specific Developmental Gene Silencing in the Honey Bee Using a Homeobox Motif. *Insect Molecular Biology* 11.July (2002): 527–532.

Beye, Martin et al. Exceptionally High Levels of Recombination Across the Honey Bee Genome. *Genome research* 16.11 (2006): 1339–44.

Bhandari, P, and L S Shashidhara. Studies on Human Colon Cancer Gene APC by Targeted Expression in Drosophila. *Oncogene* 20.47 (2001): 6871–80.

Bhanot, P et al. Frizzled and Dfrizzled-2 Function as Redundant Receptors for Wingless During Drosophila Embryonic Development. *Development* 126.18 (1999): 4175–86.

Bharathi, V et al. Genetic Characterization of the Drosophila Homologue of Coronin. *Journal of cell science* 117. (2004): 1911–22.

Biehs, B, M a Sturtevant, and E Bier. Boundaries in the Drosophila Wing Imaginal Disc Organize Vein-specific Genetic Programs. *Development* 125.21 (1998): 4245–57.

Blair, Seth S. Wing Vein Patterning in Drosophila and the Analysis of Intercellular Signaling. *Annual review of cell and developmental biology* 23 (2007): 293–319.

Bomtorin, Ana Durvalina et al. Hox Gene Expression Leads to Differential Hind Leg Development Between Honeybee Castes. *PloS one* 7.7 (2012): e40111.

Bomze, Howard M, and A Javier Lopez. Evolutionary Conservation of the Structure and Expression of Alternatively Spliced Ultrabithorax Isoforms from Drosophila. *Genetics* 136 (1994): 956– 977.

Bowsher, Julia H, Gregory A Wray, and Ehab Abouheif Ã. Growth and Patterning Are Evolutionarily Dissociated in the Vestigial Wing Discs of Workers of the Red Imported Fire Ant , Solenopsis Invicta. *Jouranl of Experimental Zoology* 776.August (2007): 769–776.

Bray, Sarah J. Notch Signalling: a Simple Pathway Becomes Complex. *Nature reviews* 7.9 (2006): 678–89.

Brisson, J a, a Ishikawa, and T Miura. Wing Development Genes of the Pea Aphid and Differential Gene Expression Between Winged and Unwinged Morphs. *Insect molecular biology* 19 Suppl 2 (2010): 63–73.

Brody, Thomas et al. Identification of Novel Drosophila Neural Precursor Genes Using a Differential Embryonic Head cDNA Screen. *Mechanisms of Development* 113 (2002): 41–59.

Brook, William J, Fernando J Diaz-benjumea, and Stephen M Cohen. Organizing Spatial Pattern in Limb Development. *Annual review of cell and developmental biology* 12 (1996): 161–80.

Brookman, J J et al. The 412 Retrotransposon and the Development of Gonadal Mesoderm in Drosophila. *Development (Cambridge, England)* 116.4 (1992): 1185–92.

Busturia, A et al. Different Forms of Ultrabithorax Proteins Generated by Alternative Splicing Are Functionally Equivalent. *The EMBO journal* 9.1 (1990): 3551–3555.

Buszczak, Michael et al. The Carnegie Protein Trap Library: a Versatile Tool for Drosophila Developmental Studies. *Genetics* 175.3 (2007): 1505–31.

Cañestro, Cristian, Hayato Yokoi, and John H Postlethwait. Evolutionary Developmental Biology and Genomics. *Nature reviews. Genetics* 8.12 (2007): 932–42.

Capdevila, Javier, and Isabel Guerrero. Targeted Expression of the Signaling Molecule Decapentaplegic Induces Pattern Duplications and Growth Alterations in Drosophila Wings. *EMBO Journal* 13.19 (1994): 4459–4468.

Capovilla, Maria, Mary Brand, and Juan Betas. Direct Regulation of Decapentaplegic by Ultrabithorax and Its Role in Drosophila M Idgut Morphogenesis. *Cell* 76 (1994): 461–475.

Carroll, Sean B. Evolution at Two Levels: On Genes and Form. *PLoS biology* 3.7 (2005): e245. ---. Homeotic Genes and the Evolution of Arthropods and Chordates. *Nature* 376 (1995): 479 – 485.

Carroll, Sean B, Scott D Weatherbee, and James A Langeland. Homeotic Genes and the Regulation and Evolution of Insect Wing Number. *Nature* 375 (1995): 58–61.

Casares, F, and E Sánchez-Herrero. Regulation of the Infraabdominal Regions of the Bithorax Complex of Drosophila by Gap Genes. *Development* 121.6 (1995): 1855–66.

Casares, Fernando et al. Interactions of Drosophila Ultrabithorax Regulatory Regions with Native and Foreign Promoters. *Genetics* 145 (1997): 123–137.

Castelli-gair, James, and Michael Akam. How the Hox Gene Ultrabithorax Specifies Two Different Segments: The Significance of Spatial and Temporal Regulation Within Metameres. *Development* 121 (1995): 2973–2982.

Castelli-Gair, James E., Jose-Luis Micol, and Antonio Garcia-Bellido. Transvection in the Drosophila Ultrabithorax Gene: A Cbx Mutant Allele Induces Ectopic Expression of a Normal Allele in Trans. *Genetics* 126 (1990): 177–184.

Cavalli, Giacomo, Valerio Orlando, and Renato Paro. Mapping DNA Target Sites of Chromatin-associated Proteins by Formaldehyde Cross-linking in Drosophila Embryos. *Chromosome Structural Analysis: A Practical Approach*. 1999. 20–37.

Choo SW, White R and Russell S . Genome wide analysis of the binding of the Hox protein Ultrabithorax and the Hox co factor Homothorax in *Drosophila*. *Plos One* 6(4). 2011

De Celis, J F, S Bray, and a Garcia-Bellido. Notch Signalling Regulates Veinlet Expression and Establishes Boundaries Between Veins and Interveins in the Drosophila Wing. *Development (Cambridge, England)* 124.10 (1997): 1919–28.

Celis, Jose F De et al. Functional Relationships Between Notch , Su (H) and the bHLH Genes of the During Imaginal Development. *Development* 122 (1996): 2719–2728.

Chang, Yuh-Long et al. A Double-bromodomain Protein, FSH-S, Activates the Homeotic Gene Ultrabithorax Through a Critical Promoter-proximal Region. *Molecular and cellular biology* 27.15 (2007): 5486–98.

Chopra, Vivek S, Joung-Woo Hong, and Michael Levine. Regulation of Hox Gene Activity by Transcriptional Elongation in Drosophila. *Current biology :* 19.8 (2009): 688–93.

Chopra, Vivek S, and Rakesh K Mishra. ‘Mir’acles in Hox Gene Regulation. *BioEssays : news and reviews in molecular, cellular and developmental biology* 28.5 (2006): 445–8.

Das, Modan K, and Ho-kwok Dai. A Survey of DNA Motif Finding Algorithms. *BMC Bioinformatics* 13 (2007):

DasGupta, Ramanuj et al. Functional Genomic Analysis of the Wnt-wingless Signaling Pathway. *Science* 308.5723 (2005): 826–33.

Dearden, Peter K. Germ Cell Development in the Honeybee (Apis Mellifera); Vasa and Nanos Expression. *BMC developmental biology* 6 (2006):

Dearden, Peter K et al. Patterns of Conservation and Change in Honey Bee Developmental Genes. *Genome research* 16.11 (2006): 1376–84.

Deshpande, N et al. The Human Transcription Enhancer Factor-1, TEF-1, Can Substitute for Drosophila Scalloped During Wingblade Development. *The Journal of biological chemistry* 272.16 (1997): 10664–8.

Duboule, Denis. The Rise and Fall of Hox Gene Clusters. *Development* 134.14 (2007): 2549–60.

Duboule, Denis, and Pascal Dolle. The Structural and Functional Organization of the Murine HOX Gene Family Resembles That of Drosophila Homeotic Genes. *The EMBO journal* 8.5 (1989): 1497–1505.

Duffy, Joseph B. GAL4 System in Drosophila : A Fly Geneticist ' s Swiss Army Knife. *Genesis* 34 (2002)

Duncan, Ian. The Bithorax Complex. *Annual review of genetics* 21 (1987): 285–319.

Estes, Patricia, Eric Fulkerson, and Yi Zhang. Identification of Motifs That Are Conserved in 12 Drosophila Species and Regulate Midline Glia Vs. Neuron Expression. *Genetics* 178.2 (2008): 787–99.

Evans, Jay D, and Daniel B Weaver. Beenome Soon: Honey Bees as a Model 'Non-model' System for Comparative Genomics. *Comparative and functional genomics* 4.3 (2003): 351–2.

Fleig, Richard, and Klaus Sander. Embryogenesis of the Honeybee Apis Mellifera L. (Hymenoptera: Apidae): A SEM Study. *International Journal of Insect morphology & embryology* 15.5 (1986): 449–462.

Foronda, David et al. Function and Specificity of Hox Genes. *The International journal of developmental biology* 53.8-10 (2009): 1404–19.

Furriols, M, and S Bray. A Model Notch Response Element Detects Suppressor of Hairless-dependent Molecular Switch. *Current biology* 11.1 (2001): 60–4.

Galant, Ron, and Sean B Carroll. Evolution of a Transcriptional Repression Domain in an Insect Hox Protein. *Nature* 415.February (2002): 910–913.

Galant, Ron, Christopher M Walsh, and Sean B Carroll. Hox Repression of a Target Gene: Extradenticle-independent, Additive Action Through Multiple Monomer Binding Sites. *Development* 129 (2002): 3115–3126.

Garaulet, Daniel L et al. Polycomb-dependent Ultrabithorax Hox Gene Silencing Induced by High Ultrabithorax Levels in Drosophila. *Development* 135.19 (2008): 3219–28.

Gavis, E R, and D S Hogness. Phosphorylation, Expression and Function of the Ultrabithorax Protein Family in Drosophila Melanogaster. *Development* 112.4 (1991): 1077–93.

Gebelein, Brian, and Richard S Mann. Compartmental Modulation of Abdominal Hox Expression by Engrailed and Sloppy-paired Patterns the Fly Ectoderm. *Developmental biology* 308.2 (2007): 593–605.

Gebelein, Brian, Daniel J Mckay, and Richard S Mann. Direct Integration of Hox and Segmentation Gene Inputs During Drosophila Development. *Nature* 431.October (2004): 653 –659.

Gehring, Walter J, Markus Affolter, and Thomas Burglin. Homeodomain Proteins. *Annual review of Biochemistry* 63 (1994): 487– 526.

Gehring, WJ, and Y Hiromi. Homeotic Genes and the Homobox. *Annual review of genetics* 20 (1986): 147– 73.

Gibson, G. Insect Evolution: Redesigning the Fruitfly. *Current biology* 9.3 (1999): 86–9.

Go, M J, D S Eastman, and S Artavanis-Tsakonas. Cell Proliferation Control by Notch Signaling in Drosophila Development. *Development (Cambridge, England)* 125.11 (1998): 2031–40.

Gong, Wei J, and Kent G Golic. Ends-out, or Replacement, Gene Targeting in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* 100.5 (2003): 2556–61.

González, Aitor, Claudine Chaouiya, and Denis Thieffry. Dynamical Analysis of the Regulatory Network Defining the Dorsal-ventral Boundary of the Drosophila Wing Imaginal Disc. *Genetics* 174.3 (2006): 1625–34.

Gould, a P, and R a White. Connectin, a Target of Homeotic Gene Control in Drosophila. *Development* 116.4 (1992): 1163–74.

Goymer, Patrick. How Old Are Hox Genes? *Nature Reviews Genetics* 8.5 (2007): 328–328.

Graba, Y, K Gieseler, et al. DWnt-4, a Novel Drosophila Wnt Gene Acts Downstream of Homeotic Complex Genes in the Visceral Mesoderm. *Development* 121.1 (1995): 209–18.

Graba, Y, D Aragnol, et al. Homeotic Control in Drosophila; the Scabrous Gene Is an in Vivo Target of Ultrabithorax Proteins. *The EMBO journal* 11.9 (1992): 3375–84.

Grenier, J K et al. Evolution of the Entire Arthropod Hox Gene Set Predated the Origin and Radiation of the Onychophoran/arthropod Clade. *Current biology* 7.8 (1997): 547–53.

Grenier, J. K., and Sean B Carroll. Functional Evolution of the Ultrabithorax Protein. *Proceedings of the National Academy of Sciences* 97.2 (2000): 704–709.

Grenier, Jennifer K, and Sean B Carroll. Functional Evolution of the Ultrabithorax Protein. *Proceedings of the National Academy of Sciences* 97 (1999): 704–709.

Harding, K, and M Levine. Gap Genes Define the Limits of Antennapedia and Bithorax Gene Expression During Early Development in Drosophila. *The EMBO journal* 7.1 (1988): 205–14.

Haring, Max et al. Chromatin Immunoprecipitation: Optimization, Quantitative Analysis and Data Normalization. *Plant methods* 3 (2007): 11.

Hatton, a R, V Subramaniam, and a J Lopez. Generation of Alternative Ultrabithorax Isoforms and Stepwise Removal of a Large Intron by Resplicing at Exon-exon Junctions. *Molecular cell* 2.6 (1998): 787–96.

Hayward, Penelope, Tibor Kalmar, and Alfonso Martinez Arias. Wnt/Notch Signalling and Information Processing During Development. *Development (Cambridge, England)* 135.3 (2008): 411–24.

Hayward, Penny et al. Notch Modulates Wnt Signalling by Associating with Armadillo/beta-catenin and Regulating Its Transcriptional Activity. *Development (Cambridge, England)* 132.8 (2005): 1819–30.

Hinz, U, a Wolk, and R Renkawitz-Pohl. Ultrabithorax Is a Regulator of Beta 3 Tubulin Expression in the Drosophila Visceral Mesoderm. *Development (Cambridge, England)* 116.3 (1992): 543–54.

Huang, Da Wei, Brad T Sherman, and Richard A Lempicki. Bioinformatics Enrichment Tools : Paths Toward the Comprehensive Functional Analysis of Large Gene Lists. *Nucleic acids research* 37.1 (2009): 1–13.

---. Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources. *Nature protocols* 4.2 (2008): 44 – 57.

Huang, Juan, Wenke Zhou, Wei Dong, et al. Directed , Efficient , and Versatile Modifications of the Drosophila Genome by Genomic Engineering. *Proceedings of the National Academy of Sciences* 106.20 (2009): 8282– 8289.

Huang, Juan, Wenke Zhou, Annie M Watson, et al. Efficient Ends-out Gene Targeting in Drosophila. *Genetics* 180.1 (2008): 703–7.

Hueber, Stefanie D et al. Comparative Analysis of Hox Downstream Genes in Drosophila. *Development* 134 (2007): 381–392.

Hueber, Stefanie D, and Ingrid Lohmann. Shaping Segments: Hox Gene Function in the Genomic Age. *BioEssays* 30.10 (2008): 965–79.

Hughes, Cynthia L, and Thomas C Kaufman. Hox Genes and the Evolution of the Arthropod Body Plan. *Evolution* 4.6 (2002): 459–499.

Huppert, S S, T L Jacobsen, and M a Muskavitch. Feedback Regulation Is Central to Delta-Notch Signalling Required for Drosophila Wing Vein Morphogenesis. *Development (Cambridge, England)* 124.17 (1997): 3283–91.

Irish, V F, a Martinez-Arias, and M Akam. Spatial Regulation of the Antennapedia and Ultrabithorax Homeotic Genes During Drosophila Early Development. *The EMBO journal* 8.5 (1989): 1527–37.

Irvine, K D et al. Negative Autoregulation by Ultrabithorax Controls the Level and Pattern of Its Expression. *Development (Cambridge, England)* 117.1 (1993): 387–99.

Irvine, K D, S L Helfand, and D S Hogness. The Large Upstream Control Region of the Drosophila Homeotic Gene Ultrabithorax. *Development (Cambridge, England)* 111.2 (1991): 407–24.

Irvine, Kenneth D, and Cordelia Rauskolb. B OUNDARIES IN D EVELOPMENT : Formation and Function. *annual reveiw of cell and Developmental Biology* 17 (2001): 189–214.

Jenner, Ronald a, and Matthew a Wills. The Choice of Model Organisms in Evo-devo. *Nature reviews. Genetics* 8.4 (2007): 311–9.

Jiang, J, and G Struhl. Regulation of the Hedgehog and Wingless Signalling Pathways by the F-box/WD40-repeat Protein Slimb. *Nature* 391.6666 (1998): 493–6.

Johnson, F B, E Parker, and M a Krasnow. Extradenticle Protein Is a Selective Cofactor for the Drosophila Homeotics: Role of the Homeodomain and YPWM Amino Acid Motif in the Interaction. *Proceedings of the National Academy of Sciences of the United States of America* 92.3 (1995): 739–43.

Kango-Singh, M, a Singh, and K P Gopinathan. The Wings of Bombyx Mori Develop from Larval Discs Exhibiting an Early Differentiated State: a Preliminary Report. *Journal of biosciences* 26.2 (2001): 167–77.

Kearney, Joseph B et al. Gene Expression Profiling of the Developing Drosophila CNS Midline Cells. *Developmental Biology* 275 (2004): 473 – 492.

Klein, Thomas. Wing Disc Development in the Fly: The Early Stages. *Current opinion in genetics & development* 11 (2001): 470–475.

Kornfeld, K et al. Structure and Expression of a Family of Ultrabithorax mRNAs Generated by Alternative Splicing and Polyadenylation in Drosophila. *Genes & Development* 3.2 (1989): 243–258. Web. 28 Feb. 2013.

Kremser, T et al. Expression of the Beta3 Tubulin Gene (beta Tub60D) in the Visceral Mesoderm of Drosophila Is Dependent on a Complex Enhancer That Binds Tinman and UBX. *Molecular & general genetics* 262.4-5 (1999): 643–58.

Laajala, Teemu D et al. A Practical Comparison of Methods for Detecting Transcription Factor Binding Sites in ChIP-seq Experiments. *BMC Genomics* 15 (2009)

Labourier, Emmanuel et al. The KH- Type RNA-binding Protein PSI Is Required for Drosophila Viability, Male Fertility, and Cellular mRNA Processing. *Genes & Development* 16 (2002): 72–84.

Langmead, Ben et al. Ultrafast and Memory-efficient Alignment of Short DNA Sequences to the Human Genome. *Genome biology* 10.3 (2009): R25.

Lecourtois, M, and F Schweisguth. The Neurogenic Suppressor of Hairless DNA-binding Protein Mediates the Transcriptional Activation of the Enhancer of Split Complex Genes Triggered by Notch Signaling. *Genes & Development* 9.21 (1995): 2598–2608.

Lemons, Derek, and William McGinnis. Genomic Evolution of Hox Gene Clusters. *Science (New York, N.Y.)* 313.5795 (2006): 1918–22.

Levine, Mike. How Insects Lose Their Limbs. *Nature* 415.6874 (2002): 848–9.

Lewis, E.B. A Gene Complex Controlling Segmentation in Drosophila. *Nature* 276 (1978): 565– 570.

Li, Heng et al. The Sequence Alignment / Map Format and SAMtools. *Bioinformatics* 25.16 (2009): 2078–2079.

Liu, Yang, Kathleen S Matthews, and Sarah E Bondos. Internal Regulatory Interactions Determine DNA Binding Specificity by a Hox Transcription Facotr. *Journal of Molecular Biology* 390.4 (2009): 760–774.

Lopez, a J, and D S Hogness. Immunochemical Dissection of the Ultrabithorax Homeoprotein Family in Drosophila Melanogaster. *Proceedings of the National Academy of Sciences of the United States of America* 88.22 (1991): 9924–8.

Maeda, Robert K, and François Karch. The ABC of the BX-C: The Bithorax Complex Explained. *Development (Cambridge, England)* 133.8 (2006): 1413–22.

Mahfooz, Najmus et al. Ubx Regulates Differential Enlargement and Diversification of Insect Hind Legs. *PloS one* 2.9 (2007): e866.

Makhijani, Kalpana et al. Modulation of Decapentaplegic Gradient During Haltere Specification in Drosophila. *Developmental biology* 302.1 (2007): 243–55.

Mallo, Moises, Deneen M Wellik, and Jacqueline Deschamps. Hox Genes and Regional Patterning of the Vertebrate Body Plan. *Developmental biology* 344.1 (2010): 7–15.

Manoonkitiwongsa, Panya S, and Robert L Schultz. Proper Nomenclature of Formaldehyde and Paraformaldehyde Fixatives for Histochemistry. *The histochemical journal* 34 (2002): 365–367.

Martinez, Pedro, and Chris T Amemiya. Genomics of the HOX Gene Cluster. *Comparative biochemistry and physiology. Part B, Biochemistry & molecular biology* 133.4 (2002): 571–80.

Massie, Charles E, and Ian G Mills. ChIPping Away at Gene Regulation. *EMBO reports* 9.4 (2008): 337–43.

Mastick, Grant S et al. Identification of Target Genes Regulated by Homeotic Proteins in Drosophila Melanogaster Through Genetic Selection of Ultrabithorax Protein Binding Sites in Yeast. *Genetics* 363 (1995): 349–363.

Masumoto, Mika, Toshinobu Yaginuma, and Teruyuki Niimi. Functional Analysis of Ultrabithorax in the Silkworm, Bombyx Mori, Using RNAi. *Development genes and evolution* 219.9-10 (2009): 437–44.

Matys, V. TRANSFAC: Transcriptional Regulation, from Patterns to Profiles. *Nucleic Acids Research* 31.1 (2003): 374–378.

McCormick, a et al. Homeotic Response Elements Are Tightly Linked to Tissue-specific Elements in a Transcriptional Enhancer of the Teashirt Gene. *Development* 121.9 (1995): 2799–812.

McGinnis, W et al. A Homologous Protein-coding Sequence in Drosophila Homeotic Genes and Its Conservation in Other Metazoans. *Cell* 37.2 (1984): 403–8.

McGregor, Alistair P et al. Morphological Evolution Through Multiple Cis-regulatory Mutations at a Single Gene. *Nature* 448.7153 (2007): 587–90.

Milner, Martin J, Alison J Bleasby, and Susan L Kelly. The Role of the Peripodial Membrane of Leg and Wing Imaginal Discs of Drosophila Melanogaster During Evagination and Differentiation in Vitro. *Roux ' s Archives of Developmental Biology* 193 (1984): 180–186.

Mockler, Todd C, and Joseph R Ecker. Applications of DNA Tiling Arrays for Whole-genome Analysis. *Genomics* 85.1 (2005): 1–15.

Mohit, Prasad et al. Modulation of AP and DV Signaling Pathways by the Homeotic Gene Ultrabithorax During Haltere Development in Drosophila. *Developmental biology* 291.2 (2006): 356–67.

Morel, Veronique et al. Transcriptional Repression by Suppressor of Hairless Involves the Binding of a Hairless-dCtBP Complex in Drosophila. *Current Biology* 11 (2001): 789–792.

Muller, Jurg, and Mariann Bienz. Ultrabithorax Expression in the Drosophila Embryo. *The EMBO journal* 10.11 (1991): 3147–3155.

De Navas, Luis F, Daniel L Garaulet, and Ernesto Sánchez-Herrero. The Ultrabithorax Hox Gene of Drosophila Controls Haltere Size by Regulating the Dpp Pathway. *Development* 133.22 (2006): 4495–506.

Nei, Masatoshi, Yoshiyuki Suzuki, and Masafumi Nozawa. The Neutral Theory of Molecular Evolution in the Genomic Era. *Annual review of genomics and human genetics* 11 (2010): 265–89.

Neto-Silva, Ricardo M, Brent S Wells, and Laura a Johnston. Mechanisms of Growth and Homeostasis in the Drosophila Wing. *Annual review of cell and developmental biology* 25 (2009): 197–220.

Neumann, C J, and S M Cohen. Long-range Action of Wingless Organizes the Dorsal-ventral Axis of the Drosophila Wing. *Development* 124.4 (1997): 871–80.

Norga, Koenraad K et al. Quantitative Analysis of Bristle Number in Drosophila Mutants Identifies Genes Involved in Neural Development. *Current Biology* 13 (2003): 1388–1396.

Ohta, Tomoko. The nearly neutral theory of molecular evolution. *Annual review of Ecol.Syst.* 23 (1992): 263–86.

Pallavi, S K, Ramakrishnan Kannan, and L S Shashidhara. Negative Regulation of Egfr/Ras Pathway by Ultrabithorax During Haltere Development in Drosophila. *Developmental biology* 296.2 (2006): 340–52.

Pallavi, S K, and L S Shashidhara. Egfr/Ras Pathway Mediates Interactions Between Peripodial and Disc Proper Cells in Drosophila Wing Discs. *Development* 130.20 (2003): 4931–41.

---. Signaling Interactions Between Squamous and Columnar Epithelia of the Drosophila Wing Disc. *Journal of cell science* 118.Pt 15 (2005): 3363–70.

Papp, Bernadett, and Jürg Müller. Histone Trimethylation and the Maintenance of Transcriptional ON and OFF States by trxG and PcG Proteins. *Genes & Development* 3 (2006): 2041–2054.

Park, Peter J. ChIP-seq: Advantages and Challenges of a Maturing Technology. *Nature reviews. Genetics* 10.10 (2009): 669–80.

Pavesi, Giulio et al. Weeder Web : Discovery of Transcription Factor Binding Sites in a Set of Sequences from Co-regulated Genes. *Nucleic acids research* 32 (2004): 199–203.

Pavlopoulos, Anastasios, and Michalis Averof. Developmental Evolution: Hox Proteins Ring the Changes. *Current Biology* 12.02 (2002): 291–293.

Pearson, Joseph C, Derek Lemons, and William McGinnis. Modulating Hox Gene Functions During Animal Body Patterning. *Nature reviews. Genetics* 6.12 (2005): 893–904.

Pen, Teresa, Isabel Rodriguez, and Juan Rafael Riesgo-escovar. A Misexpression Study Examining Dorsal Thorax Formation in Drosophila Melanogaster. *Transition* 1050.March (2002): 1035–1050.

Peter, Isabelle S, and Eric H Davidson. Evolution of Gene Regulatory Networks Controlling Body Plan Development. *Cell* 144.6 (2011): 970–85.

Petruk, Svetlana et al. Transcription of Bxd Noncoding RNAs Promoted by Trithorax Represses Ubx in Cis by Transcriptional Interference. *Cell* 127.6 (2006): 1209–21.

Piddini, Eugenia et al. Arrow (LRP6) and Frizzled2 Cooperate to Degrade Wingless in Drosophila Imaginal Discs. *Development* 132.24 (2005): 5479–89.

Poinar Jr, G.O., and B.N. Danforth. A Fossil Bee from Early Cretaceous Burmese Amber. *Science* 314.October (2006): 613.

Porcelli, Damiano et al. The Nuclear OXPHOS Genes in Insecta : a Common Evolutionary Origin , a Common Cis-regulatory Motif , a Common Destiny for Gene Duplicates. *BMC Evolutionary Biology* 17 (2007): 1–17.

Prasad, M. Regulation of Wingless and Vestigial Expression in Wing and Haltere Discs of Drosophila. *Development* 130.8 (2003): 1537–1547.

Quinlan, Aaron R, and Ira M Hall. BEDTools : a Flexible Suite of Utilities for Comparing Genomic Features. *Bioinformatics* 26.6 (2010): 841–842.

Rafel, Neus, and Marco Milán. Notch Signalling Coordinates Tissue Growth and Wing Fate Specification in Drosophila. *Development* 135.24 (2008): 3995–4001.

Reed, Hilary C et al. Alternative Splicing Modulates Drosophila melanogaster Ubx Protein Function In. (2009)

Reiter, Lawrence T et al. Proteome Comparisons Using the Negative Proteome Database ND ES RIB. *Proteins* June (2007): 164–171.

Rodríguez Dd, David Del Alamo et al. Different Mechanisms Initiate and Maintain Wingless Expression in the Drosophila Wing Hinge. *Development (Cambridge, England)* 129.17 (2002): 3995–4004.

Ronshaugen, Matthew et al. The Drosophila microRNA Iab-4 Causes a Dominant Homeotic Transformation of Halteres to Wings. *Genes & development* 19.24 (2005): 2947–52.

Ronshaugen, Matthew, Nadine Mcginnis, and William Mcginnis. Hox Protein Mutation and Macroevolution of the Insect Body Plan. *Nature* 415.February (2002): 914–917.

Rothenbuhler, Walter C. Genetics and Breeding of the Honey Bee. *Annual Review of Entomology* (1957): 161– 180.

Rubin, G M, and a C Spradling. Genetic Transformation of Drosophila with Transposable Element Vectors. *Science (New York, N.Y.)* 218 (1982): 348–53.

Ruddle, Frank H et al. Evolution of Hox Genes. *Annual review of genetics* 28 (1994): 423–442.

Saenko, Suzanne V, Marta Sp Marialva, and Patrícia Beldade. Involvement of the Conserved Hox Gene Antennapedia in the Development and Evolution of a Novel Trait. *EvoDevo* 2.1 (2011): 1–9.

Sameshima, Shin-ya, Toru Miura, and Tadao Matsumoto. Wing Disc Development During Caste Differentiation in the Ant Pheidole Megacephala (Hymenoptera : Formicidae). *Evolution & Development* 6 (2004): 336 –341.

Schulz, Cordula et al. A Misexpression Screen Reveals Effects of Bag-of-marbles and TGF Beta Class Signaling on the Drosophila Male Germ-Line Stem Cell Lineage. *Genetics* 167.June (2004): 707–723.

Sequencing, Consortium The honeybee genome. Insights into Social Insects from the Genome of the Honeybee Apis Mellifera. *Nature* 443.October (2006): 931–947.

Sharrow, Mary, and Michael Tiemeyer. Gliolectin-mediated Carbohydrate Binding at the Drosophila Midline Ensures the Fidelity of Axon Pathfinding. *Development* 128 (2001): 4585–4595.

Shashidhara, L S et al. Negative Regulation of Dorsoventral Signaling by the Homeotic Gene Ultrabithorax During Haltere Development in Drosophila. *Developmental biology* 212.2 (1999): 491–502.

Slattery, Matthew et al. Cofactor Binding Evokes Latent Differences in DNA Binding Specificity Between Hox Proteins. *Cell* 147.6 (2011): 1270–1282.

Slattery M, Ma L, Negre N, White KP and Mann R. Genome-wide tissue- specific occupancy of the Hox protein Ultrabithorax and the Hox cofactor Homothorax in *Drosophila. Plos One* 6(4) (2011).

Smith, Sheryl T et al. Genome Wide ChIP-chip Analyses Reveal Important Roles for CTCF in Drosophila Genome Organization. *Developmental biology* 328.2 (2009): 518–28.

Somorjai, Ildiko M L, and Alfonso Martinez-Arias. Wingless Signalling Alters the Levels, Subcellular Distribution and Dynamics of Armadillo and E-cadherin in Third Instar Larval Wing Imaginal Discs. *PloS one* 3.8 (2008): e2893.

Stern, Charlene A, and Tamar R Braverman. Molecular Identification , Tissue Distribution and Subcellular Localization of mST3GalV / GM3 Synthase. *Glycobiology* 10.4 (2000): 365–374.

Stern, David L. Perspective: Evolutionary Developmental Biology and the Problem of Variation. *Evolution* 54.4 (2000): 1079.

Strigini, M, and S M Cohen. Wingless Gradient Formation in the Drosophila Wing. *Current biology* 10.6 (2000): 293–300.

Struhl, Gary, and Robert A H White. Regulation of the UItrabithorax Gene of Drosophila by Other Bithorax Complex Genes. *Cell* 43.December (1985): 507–519.

Sturtevant, M a, M Roark, and E Bier. The Drosophila Rhomboid Gene Mediates the Localized Formation of Wing Veins and Interacts Genetically with Components of the EGF-R Signaling Pathway. *Genes & development* 7.6 (1993): 961–73.

Sturtevant, Mark A, and Ethan Bier. Analysis of the Genetic Hierarchy Guiding Wing Vein Development in Drosophila. *Development* 801 (1995): 785–801.

Subramaniam, Vaidyanathan, Howard M Bomze, and A Javier Lopez. Functional Differences Between Ultrabithorax Protein Isoforms. *Genetics* 136 (1994): 979– 991.

Taghli-Lamallem, Ouarda et al. Context Dependent Regulation of Hox Protein Functions by CK2 Phosphorylation Sites. *Development Genes and Evolution* 218.6 (2008): 321–332.

Tiemeyer, M, and C S Goodman. Gliolectin Is a Novel Carbohydrate-binding Protein Expressed by a Subset of Glia in the Embryonic Drosophila Nervous System. *Development* 122.3 (1996): 925–36.

Tomkins, J P et al. New Genomic Resources for the Honey bee(Apis Mellifera L.): Development of a Deep-coverage BAC Library and a Preliminary STC Database. *Genetics and molecular research : GMR* 1.4 (2002): 306–16.

Tomoyasu, Yoshinori et al. Repeated Co-options of Exoskeleton Formation During Wing-to-Elytron Evolution in Beetles. *Current Biology* 19.24 (2009): 2057–2065.

Tomoyasu, Yoshinori, Scott R Wheeler, and Robin E Denell. Ultrabithorax Is Required for Membranous Wing Identity in the Beetle Tribolium Castaneum. *Nature* 433 (2005): 643–647.

Tompa, Martin et al. Assessing Computational Tools for the Discovery of Transcription Factor Binding Sites. *Nature Biotechnology* 23.1 (2005): 137–144.

Tschopp, Patrick, and Denis Duboule. A Genetic Approach to the Transcriptional Regulation of Hox Gene Clusters. *Annual review of genetics* 45 (2011): 145–66.

Vachon, G et al. Homeotic Genes of the Bithorax Complex Repress Limb Development in the Abdomen of the Drosophila Embryo Through the Target Gene Distal-less. *Cell* 71.3 (1992): 437–50.

Vavouri, Tanya, and Greg Elgar. Prediction of Cis-regulatory Elements Using Binding Site Matrices--the Successes, the Failures and the Reasons for Both. *Current opinion in genetics & development* 15.4 (2005): 395–402.

Walldorf, U W E, and Richard Fleig. Comparison of Homeobox-containing Genes of the Honeybee and Drosophila. *Proceedings of the National Academy of Sciences* 86.December (1989): 9971–9975.

Walldorf, Uwe, Priska Binner, and Richard Fleig. Hox Genes in the Honey Bee Apis Mellifera. *Development Genes and Evolution* 210 (2000): 483–492.

Walsh, Christopher M, and Sean B Carroll. Collaboration Between Smads and a Hox Protein in Target Gene Repression. *Development* 134.20 (2007): 3585–92.

Wang, Weidong, and Gary Struhl. Drosophila Epsin Mediates a Select Endocytic Pathway That DSL Ligands Must Enter to Activate Notch. *Development* 131.21 (2004): 5367–80.

Warren, Robert W. et al. Evolution of Homeotic Gene Regulation and Function in Flies and Butterflies. *Nature* 372 (1994): 458–461.

Weatherbee, S D et al. Ultrabithorax Function in Butterfly Wings and the Evolution of Insect Wing Patterns. *Current biology : CB* 9.3 (1999): 109–15.

Weatherbee, Scott D et al. Ultrabithorax Regulates Genes at Several Levels of the Wing-patterning Hierarchy to Shape the Development of the Drosophila Haltere. *Genes & Development* 12 (1998): 1474–1482.

White, P, H Aberle, and J P Vincent. Signaling and Adhesion Activities of Mammalian Beta-catenin and Plakoglobin in Drosophila. *The Journal of cell biology* 140.1 (1998): 183–95.

Wilson, Megan J, Melanie Havler, and Peter K Dearden. Giant, Krüppel, and Caudal Act as Gap Genes with Extensive Roles in Patterning the Honeybee Embryo. *Developmental biology* 339.1 (2010): 200–11.

Zdobnov, Evgeny M, and Peer Bork. Quantification of Insect Genome Divergence. *Trends in genetics* 23.1 (1973): 16–20.

Zecca, M, K Basler, and G Struhl. Direct and Long-range Action of a Wingless Morphogen Gradient. *Cell* 87.5 (1996): 833–44.

Zhang, C C, and M Bienz. Segmental Determination in Drosophila Conferred by Hunchback (hb), a Repressor of the Homeotic Gene Ultrabithorax (Ubx). *Proceedings of the National Academy of Sciences of the United States of America* 89.16 (1992): 7511–5.

Zhang, J, and R W Carthew. Interactions Between Wingless and DFz2 During Drosophila Wing Development. *Development* 125.16 (1998): 3075–85.

Zhang, Yong et al. Model-based Analysis of ChIP-Seq (MACS). *Genome biology* 9.9 (2008): R137.