Comparative analysis of target selection by the Hox protein Ultrabithorax in *D. melanogaster* and *A. mellifera*

A thesis

Submitted in partial fulfilment of the requirements

Of the degree of

Doctor of Philosophy

By

Soumen Khan

20153387

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH, PUNE

2021

CERTIFICATE

This is to certify that the work incorporated in the thesis entitled "Comparative analysis of target selection by the Hox protein Ultrabithorax in *D. melanogaster* and *A. mellifera*" submitted by Mr. Soumen Khan 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: 31-08-2021

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

Soumen Khan
(ID Number : 20153387)

Date: 31-08-2021

Acknowledgements

I would like to start by thanking my supervisor and mentor, Prof. LS Shashidhara, who was a constant source of guidance and was always present when help was required, be it for the project or for personal and financial reasons. It has been an amazing experience to learn from someone who can always see the bigger picture and never ceases to learn.

I would also like to thank our collaborator Dr. Samir Merabet and his entire team for their inputs in this project and for hosting me at their lab. I would specifically like to thank Dr. Guillaume Giraud with whom I had numerous discussions regarding various aspects of the project as well as experiences of the PhD life.

I would like to thank my Research Advisory Committee members, Dr. Leelavati Narlikar, Prof. MS Madhusudan and Dr. Krishnapal Karmodiya for their comments and suggestions which helped in the shaping of the project. I would also like to thank Dr. Girish Ratnaparkhi for his support and for discussing and providing comments for my work.

I would like to thank Prof. Sanjeev Galande who provided resources and expertise for performing various important experiments during the course of my PhD. I would like to thank Dr. Saurabh Pradhan who performed the Library Preparation. I am grateful to Ankita Sharma for performing and teaching Cesium Chloride based isolation of plasmids, as well as for being a constant source of help throughout my PhD.

I would like to thank my lab members Dr. Jay Prakash Shukla, Pooja, Aditi and Nelchi for helping me out with dissections for ChIP-seq. I would like to thank Sanket, and Madhumita for proofreading my manuscripts and thesis and providing critical comments. I would like to thank Dilsha for helping out with fly work and for the numerous brainstorming and philosophical sessions trying to understand the intricacies of Ubx, nature and society. I would like to thank my students Garima, Vaibhav, Abhik, Pranav, Chinmay and Yash for helping me out with experiments and developing bioinformatic pipelines.

I would like to thank Mahesh Rote, Sayalee and Tushar for helping me out with administrative work. I would like to thank Mrinalini, Rupali, Kalpesh and Piyush for helping out with reagents procurement. I would like to thank Snehal and Yashwant for taking care of my fly stocks and managing with the fly food. I would like to thank Santosh and Vijay from IISER Pune Microscopy facility for training and helping me out with imagining and image analysis. I would like to thank all the founding members of the Covid-19 Testing Center (CTC)

at IISER Pune for providing us with an opportunity to engage ourselves during the difficult pandemic, especially when some of us were away from our families.

I am blessed to have some of the most wonderful and kind-hearted people as my friends in IISER. Ankita, Akanksha, Pooja, Siddhi, Sanket, Dilsha, Rajewshwari, Saurabh, Anindya, Madhumita and Vibishan are the few of the many people whom I am grateful to for making things always seem a bit easier.

I would like to thank CEFIPRA, CSIR and IISER Pune for providing funding to me as well as for the project.

Lastly, I would like to thank my family without whose support and innumerous sacrifices, none of this would have been possible.

Table of Contents

Table of Figures

List of Tables

List of Abbreviations

Ultrabithorax	Ubx
Vestigial	vg
Wingless	wg
Spalt	sal
Notch	N
Achaete	Ac
Fringe	Fng
Apterous	Ap
Scute	Sc
Sex combs reduced	Scr
quadrant enhancer of Vestigial	quad-vg
enhancer of vestigial from Apis mellifera	Apis-vg
Third thoracic segment	T3
Second Thoracic segment	T2
Thickveins	tkv
Decapentaplegic	dpp
Mothers against decapentaplegic	MAD
GAGA factor	GAF
Trithorax like	Trl
Extradenticle	Exd
Homothorax	Hth
shavenbaby	svb
Spalt major	salm
Transcription factor	TF

List of Publications

Research Articles

- **Khan S**, Pradhan SJ, Giraud G, Bleicher F, Paul R, Merabet S, Shashidhara LS. 2021. A micro-evolutionary change in target binding sites as key determinant of Ultrabithorax function in *Drosophila*; *bioRxiv* 2021.08.17.456507. doi:10.1101/2021.08.17.456507

- Paul R, Giraud G, Domsch K, **Khan S** et al. Hox dosage contributes to flight appendage morphology in *Drosophila*. *Nat Commun*. 2021;12(1). doi:10.1038/s41467-021-23293-8

- Giraud, G., Paul, R., Duffraisse, M., **Khan, S.**, Shashidhara, L. S., & Merabet, S. (2021). Developmental Robustness: The Haltere Case in *Drosophila*. *Frontiers in cell and developmental biology*, *9*, 713282. https://doi.org/10.3389/fcell.2021.713282

- **Khan S**, Bhattacharjee A, Badhe P, Dongaonkar C, Gaglani Y, Giraud G, Merabet S, Shashidhara LS. An automated module for parsing ChIP and RNA sequencing combined with gene regulatory pattern predictions and its applications in predicting novel cofactors for Ubx in the context of haltere specification (manuscript in preparation).

Review Articles

- **Khan S**, Dilsha C, Shashidhara LS. Haltere development in *D. melanogaste*r: Implications for the evolution of appendage size, shape and function. *Int J Dev Biol*. 2020;64(1-3):169-175. doi:10.1387/ijdb.190133LS

Synopsis

Topic: Target selection by the Hox protein Ultrabithorax in *Drosophila melanogaster* halteres and its implications in evolution of hindwing morphology in insects

Name of the student: Soumen Khan
Name of the Thesis advisor: LS Shashidhara

Indian Institute of Science Education and Research, Pune

Introduction

The diversification of body plans and the emergent functions of appendages have played a critical role in the evolution of the metazoan lineage. This is best exemplified in the diversity of appendage structures observed amongst closely related arthropod species. In Bilaterans, Hox genes regulate the development of serially homologous structures along the anterior posterior axis(Lewis, 1978; McGinnis and Krumlauf, 1992). Variations in the expression, regulation and function of Hox proteins, thus, seem to be important for morphological diversification. However, all Hox proteins are known to bind, in-vitro, with low affinity and specificity to a motif containing a TAAT core site (Berger et al., 2008; Ekker et al., 1991; Noyes et al., 2008). Thus, the mechanisms governing body plan diversification by Hox protein across different species still remain elusive.

Various attempts have been made in the past two decades to understand how Hox proteins select their targets. Among these, candidate gene approach has established the role of multiple monomeric Hox sites (Galant et al., 2002) and showed importance of low affinity binding sites for specificity and robustness of Hox protein function (Crocker et al., 2015). Similarly, role of cofactors of Hox proteins such as the TALE class of cofactors, particularly PBC and Meis family proteins, in regulation of Hox target genes has been demonstrated (Mann and Affolter, 1998; reviewed in Merabet and Galliot, 2015). However, various distal appendages in arthropods and vertebrates do not require the input of either the PBC or Meis

family of cofactors which might be suggestive of a cofactor independent mechanism of target selection.

In *Drosophila melanogaster*, the Hox protein Ultrabithorax specifies the development of halteres in the third thoracic (T3) segment. Loss of function mutations of Ubx give rise to four winged flies with a complete duplication of the T2 segment in place of T3 (Lewis, 1978), while overexpression of Ubx in larval wing discs lead to wing to haltere transformations (Cabrera et al., 1985; Castelli-Gair et al., 1990; White and Akam, 1985; White and Wilcox, 1985). Like all Hox proteins, Ubx is also known to bind in-vitro to a motif with a consensus sequence "TTAATKR". However, whole genome studies in the *Drosophila* halteres do not show enrichment for this motif (Agrawal et al., 2011; Choo et al., 2011), indicating that in-vivo, this motif might not be the preferred motif for Ubx binding. While motifs for other transcription factor like GAGA factor and MAD seem to be enriched in Ubx pulled down sequences, a thorough understanding of how Ubx selects its targets in *Drosophila* halteres is still lacking.

Interestingly, Ubx is also expressed in hindwing primordia of different insect species (Carroll 1995; Prasad, Tarikere et al. 2016). In Coleopterans like *Tribolium castaneum*, Ubx is expressed and required for the proper specification of hindwings in T3 (Tomoyasu et al., 2005). Similarly, in Lepidopterans like *Precis coenia*, Ubx is required for specifying differences in eyespot patterns between the forewing and hindwings (Weatherbee et al., 1999). In Hymenopterans, like *Apis mellifera*, Ubx is expressed in both forewing and hindwing primordia, with the expression being marginally stronger in the hindwing primordia (Prasad and Tarikere et al., 2016). Comparison of protein structure of Ubx across different insect species indicate high level of conservation at the DNA binding domain. Studies carried out by Prasad and Tarikere et al., 2016 reveal that a large number of genes have remained common targets of Ubx between *Drosophila melanogaster* and *Apis mellifera*. Amongst these, a large proportion of genes were involved in wing patterning pathways. Interestingly, some of these common targets like *vestigial (vg)* and *wingless (wg)* are differentially regulated between the T2 and T3 appendage primordia in *Drosophila* and *Apis*, suggesting that differences in hindwing morphology between the two species is a characteristic of differences in regulation of common target genes. Even though comparison of enhancer sequences of the *vestigial* gene from *Drosophila* and *Apis* indicate a difference in binding motif of the Adf1 transcription factor, genetic experiments using mutant enhancers have so far remained inconclusive.

In this study, we have used a genome wide strategy to identify recognition sequences that are bound by Ubx in *Drosophila* halteres. We find that a motif with a TAAAT core sequence is specifically enriched by Ubx and is bound with high affinity and specificity by Ubx in-vitro. This motif was functionally critical for the activation of a target gene of Ubx, *CG13222*, in *Drosophila* halteres. Additionally, the presence of this motif was sufficient to bring an otherwise unresponsive enhancer of vestigial gene from *Apis* under the control of Ubx in transgenic *Drosophila*. Our studies further indicate that microevolutionary changes in binding motif of Ubx between *Apis* and *Drosophila* might have brought important wing patterning genes under the regulation of Ubx in the dipteran lineage, thereby facilitating haltere specification.

Specific Objectives of the Study

1. To identify in-vivo binding motifs for the Ubx protein in *Drosophila melanogaster* halteres using ChIP sequencing
2. To elucidate the functional role of a strong binding motif for Ubx containing a TAAAT core in regulation of target genes in *Drosophila* halteres
3. To understand the role of the strong binding Ubx motif containing a TAAAT core in regulation of an enhancer of the *vestigial* gene from *Apis mellifera* in transgenic *Drosophila*
4. To compare mechanisms of target selection of the Ubx protein between *Drosophila* halteres and *Apis* hindwings
5. To understand the mechanisms governing recognition of activated vs repressed genes by Ubx

Results

1. Motif with a TAAAT core sequences is enriched by Ubx in *Drosophila* halteres

We performed genome wide analysis of Ubx binding in *Drosophila* halteres using Chromatin immunoprecipitation coupled to massively parallel sequencing (ChIP-seq). Using three biological replicates, our study identified 2171 high confidence peaks which corresponded to 1236 direct targets. We performed denovo motif analysis of Ubx pulled down sequences and identified a motif containing a TAAAT core sequence (hereafter referred to as the TAAAT motif) to be enriched by Ubx. The frequency of the TAAAT motif was 1.7 folds higher in Ubx pulled down sequences as compared to the entire genome, suggesting that it is the preferred binding motif for Ubx in *Drosophila* halteres. We further analyzed the frequency of the

TAAAT motif in enhancers of genes that are upregulated and downregulated and found them to be quite similar. This indicates that while the TAAAT motif might be used for target selection, it might not help in recognition of activated vs repressed targets by Ubx in *Drosophila* halteres.

2. The TAAAT motif is critical for Ubx mediated activation of *CG13222*

We sought to understand the functional role of the TAAAT motif in the context of Ubx mediated haltere specification using well characterized targets of Ubx. The *CG13222* gene is upregulated by Ubx in *Drosophila* halteres. Previous studies conducted by Hersh et al., 2007 identified two Ubx binding sites containing a TAAT core sequence (hereafter referred to as TAAT motif) in the *edge* enhancer of *CG13222*, which they termed as site1 and site2. While the site1 was critical for the expression of the enhancer, the site2 was dispensable. We identified that the site1 of the edge enhancer had a TAAAT motif in continuation with the TAAT motif. The TAAAT motif was not only necessary for binding of Ubx to the site1 of the edge enhancer in-vitro, but also critical for the Ubx mediated activation of the edge enhancer in haltere imaginal discs as well as S2 cells. Additionally, changing the TAAT motif in the site2 of the enhancer to TAAAT motif brought the enhancer under the added control of Ubx. Our results, thus, suggest a critical role of the TAAAT motif in the Ubx mediated activation of the *CG13222* gene.

3. The TAAAT motif is sufficient to bring an otherwise unresponsive enhancer of the *vestigial* gene from *Apis* in *Drosophila* transgenics

To understand the role of the TAAAT motif in Ubx mediated downregulation of target genes, we studied enhancers of Ubx targets that are downregulated in the halteres. We find that the quadrant enhancer of the *vestigial* gene (hereafter referred to as *quad-vg*) had both TAAT and TAAAT motifs. This was particularly interesting since an orthologous enhancer of the *vestigial* gene from *Apis mellifera* (hereafter referred to as *Apis-vg*), which shows similar expression patterns in both wing and haltere imaginal discs in *Drosophila* transgenics, had only a TAAT motif and no TAAAT motif. To understand the role of TAAAT motif in regulation of *vg* in halteres, we generated transgenic *Drosophila* carrying mutations in the TAAAT motif in the *quad-vg* enhancer. However, our results were inconclusive as most mutations in the TAAAT or TAAT motifs led to complete loss of reporter expression in both wing and haltere imaginal discs. We next generated *Drosophila* transgenics carrying mutations in the *Apis-vg* enhancer.

Our results reveal that replacing the TAAT motif with the TAAAT motif in the *Apis-vg* enhancer was sufficient to bring the otherwise unresponsive *Apis-vg* enhancer under the control of Ubx. Additionally, we observe that the repression of the *Apis-vg* enhancer was specific to the pouch region, suggesting that changing the TAAT to TAAAT motif makes the *Apis-vg* enhancer's expression closer to the *quad-vg* enhancer. Our results, thus, suggest that microevolutionary changes in Ubx binding motifs (TAAT to TAAAT) might have brought enhancers of important wing patterning genes like *vestigial* under the regulation of Ubx in the dipteran lineage.

4. The TAAAT motif is enriched specifically in enhancers of genes from *Drosophila melanogaster* but not in *Apis mellifera*

To understand whether the difference in Ubx binding motifs between *Drosophila* and *Apis* are specific to the *vestigial* gene or are a recurrent theme impinging on differences in hindwing morphology, we compared genome wide binding profiles for Ubx in the two species. Comparison of Ubx protein sequences between *Drosophila, Apis* as well as other insects indicate strong level of conservation at the DNA binding homeodomain region. Similar to previous studies, our results also reveal that wing patterning genes are common targets of Ubx in both species. However, while denovo motif analysis of Ubx pulled down sequences in *Drosophila* halteres reveal enrichment of the TAAAT motif, we did not find any such enrichment in *Apis* hindwings. Additionally, the frequency of the TAAAT motif in the *Apis* genome was 1.8times higher than that observed for the *Drosophila* genome. Our results, thus, suggest that the TAAAT motif is selectively enriched in enhancers of targets in *Drosophila* and thus might be involved in Ubx mediated target selection specifically in the dipteran lineage.

5. Mechanisms governing recognition of activated vs repressed genes by Ubx

While our studies suggest the role of the TAAAT motif in Ubx mediated target selection in *Drosophila* halteres, its importance in upregulation as well as downregulation of target genes indicate that it is not involved in differentiating between activated vs repressed targets by Ubx. We tried to understand the molecular cues or transcriptional codes that are critical for Ubx to recognize which genes to upregulate and which to downregulated using an in-silico approach. We hypothesized that transcription factor profile differences between enhancers of genes that are upregulated and downregulated can provide critical insights into the transcriptional codes identified by Ubx. We first compared Ubx direct targets to differentially expressed genes

between wing and haltere imaginal discs and identified Ubx bound enhancers corresponding to genes that are upregulated and downregulated in the haltere. We then developed a GUI based analysis pipeline to scan for transcription factors in each of the enhancer categories. Our results reveal that a motif predicted to be bound by the Mes2 protein was specifically enriched in enhancers of genes that are upregulated by Ubx. We find that the binding motif for Mes2 was present in the enhancer of the *CG13222* gene which is upregulated by Ubx in halteres and was required for its Ubx mediated regulation in S2 cells. However, we did not find substantial evidence for the role of Mes2 in activation of the *CG13222* enhancer in haltere imaginal discs. Taken together, while our efforts help throw some light into transcription factor profiles that might be relevant for discriminating between upregulated and downregulated targets by Ubx in *Drosophila* halteres, these need to be extensively and empirically tested. Perhaps, technological advances involving machine learning algorithms might help solve the mystery as to how Hox proteins like Ubx recognize targets.

References

Agrawal P, Habib F, Yelagandula R, Shashidhara LS. 2011. Genome-level identification of targets of Hox protein Ultrabithorax in Drosophila: novel mechanisms for target selection. *Sci Rep* **1**:1–10. doi:10.1038/srep00205

Appel B, Sakonju S. 1993. Cell-type-specific mechanisms of transcriptional repression by the homeotic gene products UBX and ABD-A in Drosophila embryos. *EMBO J* **12**:1099–1109.

Aza-Blanc P, Kornberg TB. 1999. Ci: a complex transducer of the hedgehog signal. *Trends Genet* **15**:458–462. doi:10.1016/s0168-9525(99)01869-7

Bender W, Akam M, Karch F, Beachy PA, Peifer M, Spierer P, Lewis EB, Hogness DS. 2007. Molecular genetics of the bithorax complex in Drosophila melanogaster. *Genes, Dev Cancer Life Work Edward B Lewis* **2**:311–326. doi:10.1007/978-1-4020-6345-9_14

Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Peña-Castillo L, Alleyne TM, Mnaimneh S, Botvinnik OB, Chan ET, Khalid F, Zhang W, Newburger D, Jaeger SA, Morris QD, Bulyk ML, Hughes TR. 2008. Variation in Homeodomain DNA Binding Revealed by High-Resolution Analysis of Sequence Preferences. *Cell* **133**:1266–1276. doi:10.1016/j.cell.2008.05.024

Brook WJ, Diaz-Benjumea FJ, Cohen SM. 1996. Organizing Spatial Pattern in Limb Development. *Annu Rev Cell Dev Biol* **12**:161–180.

doi:10.1146/annurev.cellbio.12.1.161

Cabrera C V., Botas J, Garcia-Bellido A. 1985. Distribution of Ultrabithorax proteins in mutants of Drosophila bithorax complex and its transregulatory genes. *Nature* **318**:569–571. doi:10.1038/318569a0

Carroll SB. 1995. Homeotic genes and the evolution of arthropods and chordates. *Nature* **376**:479–485. doi:10.1038/376479a0

Carroll SB, Weatherbee SD, Langeland JA. 1995. Homeotic genes and the regulation and evolution of insect wing number. *Nature* **375**:58–61. doi:10.1038/375058a0

Carroll Sean B. , Grenier Jennifer K. WSD. 2004. From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design.

Castelli-Gair JE, Micol JL, Garcia-Bellido A. 1990. Transvection in the Drosophila Ultrabithorax gene: A Cbx1 mutant allele induces ectopic expression of a normal allele in trans. *Genetics* **126**:177–184. doi:10.1093/genetics/126.1.177

Choo SW, White R, Russell S. 2011. Genome-wide analysis of the binding of the hox protein ultrabithorax and the hox cofactor homothorax in Drosophila. *PLoS One* **6**:25–27. doi:10.1371/journal.pone.0014778

Crickmore MA, Mann RS. 2007. Hox control of morphogen mobility and organ development through regulation of glypican expression. *Development* **134**:327–334. doi:10.1242/dev.02737

Crickmore MA, Mann RS. 2006. Hox control of organ size by regulation of morphogen production and mobility. *Science (80-)* **313**:63–68. doi:10.1126/science.1128650

Crocker J, Abe N, Rinaldi L, McGregor AP, Frankel N, Wang S, Alsawadi A, Valenti P, Plaza S, Payre F, Mann RS, Stern DL. 2015. Low Affinity Binding Site Clusters Confer Hox Specificity and Regulatory Robustness. *Cell* **160**:191–203. doi:https://doi.org/10.1016/j.cell.2014.11.041

Das PM, Ramachandran K, Singal R. 2004. Das376.indd **37**:1–9.

Davidson EH. 2001. Genomic Regulatory Systems. doi:https://doi.org/10.1016/B978-0-12-205351-1.X5000-9

De Las Heras JM, García-Cortés C, Foronda D, Pastor-Pareja JC, Shashidhara LS, Sánchez-Herrero E. 2018. The Drosophila Hox gene Ultrabithorax controls appendage shape by regulating extracellular matrix dynamics. *Dev* **145**:1–10. doi:10.1242/dev.161844

De Navas LF, Reed H, Akam M, Barrio R, Alonso CR, Sánchez-Herrero E. 2011. Integration of RNA processing and expression level control modulates the function of the

Drosophila Hox gene Ultrabithorax during adult development. *Development* **138**:107–116. doi:10.1242/dev.051409

Diaz-Benjumea FJ, Cohen SM. 1993. Interaction between dorsal and ventral cells in the imaginal disc directs wing development in Drosophila. *Cell* **75**:741–752. doi:10.1016/0092-8674(93)90494-b

Ekker SC, Young KE, Von Kessler DP, Beachy PA. 1991. Optimal DNA sequence recognition by the Ultrabithorax homeodomain of Drosophila. *EMBO J* **10**:1179–1186. doi:10.1002/j.1460-2075.1991.tb08058.x

Francois J. 1977. Evolution and Tinkering **196**:1161–1166.

Galant R, Walsh CM, Carroll SB. 2002. Hox repression of a target gene: extradenticle-independent, additive action through multiple monomer binding sites. *Development* **129**:3115–3126.

Hersh BM, Carroll SB. 2005. Direct regulation of knot gene expression by Ultrabithorax and the evolution of cis-regulatory elements in Drosophila. *Development* **132**:1567–1577. doi:10.1242/dev.01737

Hersh BM, Nelson CE, Stoll SJ, Norton JE, Albert TJ, Carroll SB. 2007. The UBX-regulated network in the haltere imaginal disc of D . melanogaster **302**:717–727. doi:10.1016/j.ydbio.2006.11.011

Hughes CL, Kaufman TC. 2002. Hox genes and the evolution of the arthropod body plan. *Evol Dev* **4**:459–499. doi:10.1046/j.1525-142x.2002.02034.x

Irvine KD, Harvey KF. 2015. Control of organ growth by patterning and hippo signaling in drosophila. *Cold Spring Harb Perspect Biol* **7**:1–16. doi:10.1101/cshperspect.a019224

Johnston DS, Nüsslein-Volhard C. 1992. The origin of pattern and polarity in the Drosophila embryo. *Cell* **68**:201–219. doi:10.1016/0092-8674(92)90466-P

Kim J, Sebring A, Esch JJ, Kraus ME, Vorwerk K, Magee J, Carroll SB. 1996. Integration of positional signals and regulation of wing formation and identity by Drosophila vestigial gene. *Nature* **382**:133–138. doi:10.1038/382133a0

Lappin TRJ, Grier DG, Thompson A, Halliday HL. 2006. HOX genes: Seductive science, mysterious mechanisms. *Ulster Med J* **75**:23–31.

Lewis EB. 1978. A gene complex controlling segmentation in Drosophila. *Genes, Dev Cancer Life Work Edward B Lewis* 229–242. doi:10.1007/978-1-4020-6345-9_10

Makhijani K, Kalyani C, Srividya T, Shashidhara LS. 2007. Modulation of Decapentaplegic gradient during haltere specification in Drosophila. *Dev Biol* **302**:243–255.

doi:10.1016/j.ydbio.2006.09.029

Mann RS, Affolter M. 1998. Hox proteins meet more partners. *Curr Opin Genet Dev* **8**:423–429. doi:10.1016/s0959-437x(98)80113-5

McGinnis W, Krumlauf R. 1992. Homeobox genes and axial patterning. *Cell* **68**:283–302. doi:10.1016/0092-8674(92)90471-N

Merabet S, Galliot B. 2015. The TALE face of Hox proteins in animal evolution. *Front Genet* **6**:1–13. doi:10.3389/fgene.2015.00267

Mohit P, Makhijani K, Madhavi MB, Bharathi V, Lal A, Sirdesai G, Reddy VR, Ramesh P, Kannan R, Dhawan J, Shashidhara LS. 2006. Modulation of AP and DV signaling pathways by the homeotic gene Ultrabithorax during haltere development in Drosophila **291**:356–367. doi:10.1016/j.ydbio.2005.12.022

Morris SC. 1993. The fossil record and the early evolution of the Metazoa. *Nature* **361**:219–225. doi:10.1038/361219a0

Noyes MB, Christensen RG, Wakabayashi A, Stormo GD, Brodsky MH, Wolfe SA. 2008. Analysis of Homeodomain Specificities Allows the Family-wide Prediction of Preferred Recognition Sites. *Cell* **133**:1277–1289. doi:10.1016/j.cell.2008.05.023

Ohno S. 1970. Evolution by Gene Duplication. doi:10.1007/978-3-642-86659-3

Pallavi SK, Shashidhara LS. 2003. Egfr/Ras pathway mediates interactions between peripodial and disc proper cells in Drosophila wing discs. *Development* **130**:4931–4941. doi:10.1242/dev.00719

Passner JM, Ryoo HD, Shen L, Mann RS, Aggarwal AK. 1999. Structure of a DNA-bound Ultrabithorax ± Extradenticle homeodomain complex. *Nature* **397**:714–719.

Pearson JC, Lemons D, McGinnis W. 2005. Modulating Hox gene functions during animal body patterning. *Nat Rev Genet* **6**:893–904. doi:10.1038/nrg1726

Prasad N, Tarikere S, Khanale D, Habib F, Shashidhara LS. 2016a. A comparative genomic analysis of targets of Hox protein Ultrabithorax amongst distant insect species. *Sci Rep* **6**:27885. doi:10.1038/srep27885

Prasad N, Tarikere S, Khanale D, Habib F, Shashidhara LS. 2016b. A comparative genomic analysis of targets of Hox protein Ultrabithorax amongst distant insect species. *Sci Rep* **6**:27885. doi:10.1038/srep27885

Prince VE. 2002. The Hox paradox: More complex(es) than imagined. *Dev Biol* **249**:1–15. doi:10.1006/dbio.2002.0745

Roch F, Akam M. 2000. Ultrabithorax and the control of cell morphology in Drosophila

halteres. *Development* **127**:97–107.

Rogulja D, Irvine KD. 2005. Regulation of cell proliferation by a morphogen gradient. *Cell* **123**:449–461. doi:10.1016/j.cell.2005.08.030

Ruiz-Losada M, Blom-Dahl D, Córdoba S, Estella C. 2018. Specification and patterning of Drosophila appendages. *J Dev Biol* **6**. doi:10.3390/jdb6030017

Shashidhara LS, Agrawal N, Bajpai R, Bharathi V, Sinha P. 1999. Negative regulation of dorsoventral signaling by the homeotic gene Ultrabithorax during haltere development in Drosophila. *Dev Biol* **212**:491–502. doi:10.1006/dbio.1999.9341

Singh S, Sánchez-Herrero E, Shashidhara LS. 2015. Critical role for Fat/Hippo and IIS/Akt pathways downstream of Ultrabithorax during haltere specification in Drosophila. *Mech Dev* **138**:198–209. doi:10.1016/j.mod.2015.07.017

Slattery M, Ma L, Négre N, White KP, Mann RS. 2011a. Genome-wide tissue-specific occupancy of the hox protein ultrabithorax and hox cofactor homothorax in Drosophila. *PLoS One* **6**. doi:10.1371/journal.pone.0014686

Slattery M, Riley T, Liu P, Abe N, Gomez-Alcala P, Dror I, Zhou T, Rohs R, Honig B, Bussemaker HJ, Mann RS. 2011b. Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* **147**:1270–1282. doi:10.1016/j.cell.2011.10.053

Struhl G. 1981. A homoeotic mutation transforming leg to antenna in Drosophila. *Nature* **292**:635–638. doi:10.1038/292635a0

Tomoyasu Y, Wheeler SR, Denell RE. 2005. Ultrabithorax is required for membranous wing identity in the beetle Tribolium castaneum. *Nature* **433**:643–647. doi:10.1038/nature03272

Walsh Christopher M., Carroll SB. 2007. Collaboration between Smads and a Hox protein in target gene repression. *Development* **134**:3585–3592. doi:10.1242/dev.009522

Walsh C. M., Carroll SB. 2007. Collaboration between Smads and a Hox protein in target gene repression. *Development* **134**:3585–3592. doi:10.1242/dev.009522

Weatherbee SD, Halder G, Kim J, Hudson A, Carroll S. 1998. Ultrabithorax regulates genes at several levels of the wing-patterning hierarchy to shape the development of the Drosophila haltere. *Genes Dev* **12**:1474–1482. doi:10.1101/gad.12.10.1474

Weatherbee SD, Nijhout HF, Grunert LW, Halder G, Galant R, Selegue J, Carroll S. 1999. Uitrabithorax function in butterfly wings and the evolution of insect wing patterns. *Curr Biol* **9**:109–115. doi:10.1016/S0960-9822(99)80064-5

White RAH, Akam ME. 1985. Contrabithorax mutations cause inappropriate expression of Ultrabithorax products in Drosophila. *Nature* **318**:567–569. doi:10.1038/318567a0

White RAH, Wilcox M. 1985. Regulation of the distribution of Ultrabithorax proteins in Drosophila. *Nature* **318**:563–567. doi:10.1038/318563a0

Chapter 1

Introduction

The diversity of form and function has been a critical step of the evolution of the metazoan lineages. This is best portrayed for one of the most successful phyla on earth- arthropods, which display huge diversity in body plans suitable for the adaptation of their environment (Carroll, 1995). Remarkably though, there appears to be a shared blueprint which has undergone numerous innovations and modifications over more than 500 million years of evolution. (Morris, 1993).

So, what forms the basis for the evolution of body parts and body plans? While it is apparent that differences in the genetic makeup impinge on the final morphology of an organism, the mechanisms employed for the differential development of appendages in different organisms, like limbs in mammals or wings in insects, seem to be built upon a similar logic.

Early studies focusing on the nature of genetic evolution postulated two main mechanisms: a) duplication and divergence of genes (Ohno, 1970) and b) regulatory changes in gene expression (Francois, 1977). However, to understand how duplication, divergence as well as regulatory differences might affect morphological diversification, it was important to identify the genes which regulate morphology. The discovery of Hox genes, thus, was one of the most significant events for understanding the mechanisms of morphological evolution. The translation product of Hox genes are transcription factors which were identified to be master regulators involved in the development of serially homologous structures along the antero-posterior axis (Lewis, 1978; McGinnis and Krumlauf, 1992). Hox proteins were found to be present in most metazoan species and were implicated in morphological diversification of metazoan species (reviewed in Carroll, 1995).

Investigations into the mechanisms of regulation by Hox proteins presented with one of the most intriguing paradox. On one hand, Hox proteins are highly conserved in terms of their DNA binding domains and are known to bind with similar affinity and specificity to a motif with a TAAT core in-vitro (Berger et al., 2008; Ekker et al., 1991; Noyes et al., 2008). On the other hand, Hox proteins display remarkable specificity of regulation and function by activating as well as repressing a large number of genes in-vivo. This paradoxical nature of Hox specificity has eluded developmental biologists for many decades. Several approaches have been employed in past two decades to solve this paradox using single gene approaches (Galant et al., 2002; Slattery et al., 2011) as well as genome wide studies (Agrawal et al., 2011; Choo et al., 2011; Slattery et al., 2011). Cooperation with other transcription factors

(Christopher M. Walsh and Carroll, 2007) as well as binding to low affinity homotypic clusters (Crocker et al., 2015) are two of the most accepted mechanisms which helped throw some light on the specificity of target regulation by Hox genes. However, the mechanistic cues demarcating the segment specific activation or repression of a target by a Hox gene, especially when Hox genes are expressed in overlapping segments, remains largely answered. Additionally, with a core TAAT motif as the only recognition motif, the mechanisms by which Hox genes recognize their targets are also not well understood.

Another intriguing feature of Hox proteins, famously known as the Hox paradox, is the fact that orthologous Hox genes in different organisms specify different appendages (reviewed in Prince, 2002). For example, in the dipteran *Drosophila melanogaster*, the Hox protein Ultrabithorax (Ubx) specifies a modified wing like structure, known as halteres, in the third thoracic segment (T3) (Lewis, 1978). On the other hand, in the coleopteran *Tribolium castaneum*, Ubx is necessary for the specification of hindwings in the T3 (Tomoyasu et al., 2005). Comparison of Ubx protein sequences between *Drosophila* and *Tribolium* as well as other insect species, however, reveal high degree of conservation within the DNA binding domain, thereby suggesting that differences in morphology between the two species is not due to changes in the Ubx protein per se. Rather, it has been reported that the morphological differences specified by homologous Hox proteins are a consequence of differential regulation of common targets and not in differences in the targets themselves (Davidson, 2001; Carroll Sean B., Grenier Jennifer K., 2004). Thus, understanding how Hox genes select and regulate their targets and how such regulations change over the course of evolution is key to understanding mechanisms of morphological diversification.

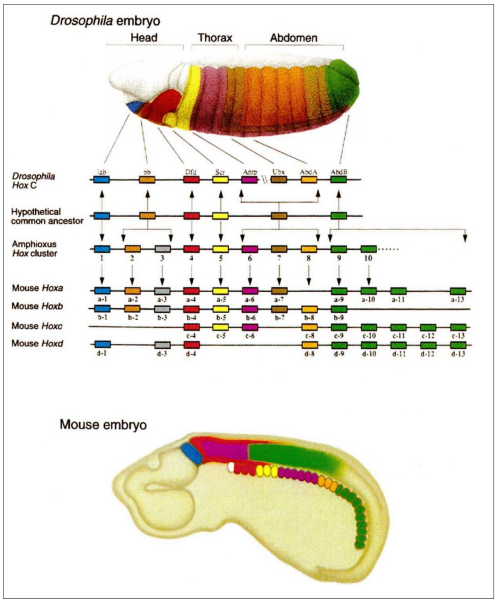
In the sections that follow, we provide a brief introduction of Hox genes, specifically the Ultrabithorax gene in *Drosophila melanogaster* and its role in the specification of halteres. We further elaborate the mechanisms of target gene selection and regulation known for the Ubx protein in the context of haltere specification as well as in the context of evolution of hindwing morphology in different insect species. Finally, we outline the specific objectives of the study and provide an account of the studies performed to address those objectives.

1.1. Hox Genes

The development of a multicellular organism from a single cell is one of the most fascinating events in biological systems. The final structure of an organism is highly complex involving varied cell types and their organization in a highly specific yet elaborate plan along a defined axis. However, the positional information required for specifying such complex structures are determined at the embryonic level.

Using *Drosophila melanogaster* as the model system, various studies in the last century have identified maternal effector genes which help to establish morphogen gradients post fertilization (reviewed in Johnston and Nüsslein-Volhard, 1992). The protein products of maternal effector genes diffuse freely and establish positional information in the zygote. These regulatory proteins activate or repress the first group of zygotic genes known as gap genes. The varying combinations and concentrations of gap genes, which encode transcription factors, regulate the transcription of pair-rule genes which in turn activate the segment-polarity genes, thus diving the embryo into 14 segment wise units. This regulatory cascade also controls the expression of Homeotic selector genes which specify the development of each segment along the antero-posterior axis.

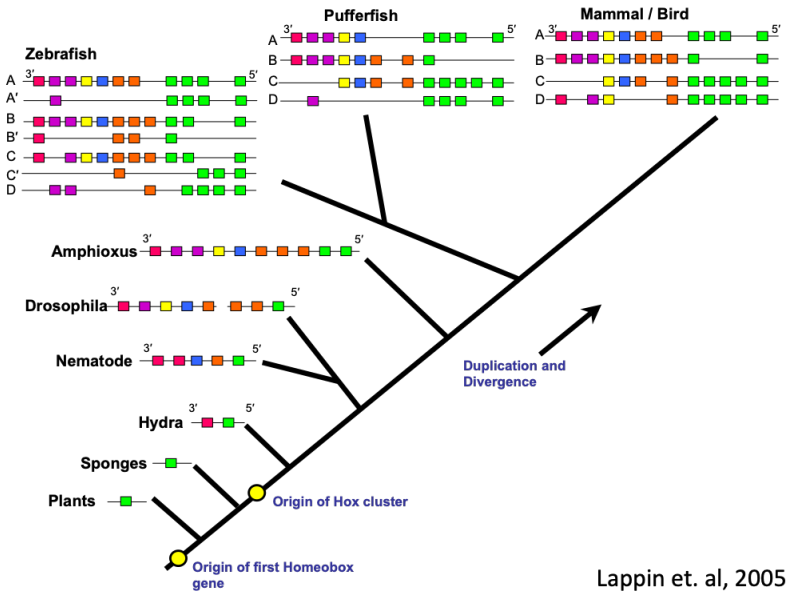
Homeotic selector genes or Hox genes encode a group of transcription factors which are expressed in a segment-specific manner along the anterior-posterior (AP) axis of an organism and specify the development of that segment (Lewis, 1978; McGinnis and Krumlauf, 1992; Struhl, 1981). These proteins contain a 60 amino acid long DNA binding domain, known as the homeodomain, which is highly conserved across all metazoan species (reviewed in Lappin et al., 2006). The role of Hox genes in animal development is evident from the fact that any ectopic expression of these genes leads to severe transformation of one body part to other (homeotic transformations) (Lewis, 1978; Struhl, 1981). Additionally, the variation in number, expression and function of Hox genes has been directly correlated to the diversification of body plans along the anterior posterior axis, suggesting their critical role in the metazoan evolution (reviewed in Carroll, 1995; Hughes and Kaufman, 2002; Pearson et al., 2005).



Carroll, 1995

Fig 1.1: Hox genes and development of an organism along the AP axis

The topmost panel shows Hox gene organization in the Drosophila embryo. The middle panel shows the relation between Hox clusters in different metazoan species. The bottom panel shows Hox gene organization in the mouse embryo. Hox genes are conserved in terms of their expression domains



Lappin et. al, 2005

Fig 1.2: Hox gene duplications and divergence

The duplication and divergence of Hox clusters have been important events in the body plan diversification in animal evolution.

Studies investigating the function and mechanism of regulation by Hox genes have primarily been done using the *Drosophila melanogaster* as the model system, which contains 8 Hox genes that are expressed along the AP axis (reviewed in Pearson et al., 2005). The flight appendages in *Drosophila* consist of a pair of wings in the second thoracic segment (T2) and a pair of modified wing structures known as the halteres in the third thoracic segment (T3). The wing is still considered to be a Hox free state. The halteres, on the other hand, are specified by the Hox protein Ultrabithorax (Ubx). Loss of Ubx function in developing halteres lead to haltere to wing transformations (Lewis, 1978). Conversely, overexpression of Ubx in developing wings lead to severe wing to haltere transformations (Cabrera et al., 1985; Castelli-Gair et al., 1990; White and Akam, 1985; White and Wilcox, 1985), thus suggesting that Ubx is required and sufficient for the haltere specification process. The differential development of the wing and halteres in *Drosophila melanogaster*, thus, form a good system to understand Hox mediated cell fate determination. Additionally, the *Drosophila* wing and haltere system represent the evolutionary events that have established differences in serially homologous structures like fore and hind wings in insects, wings and legs in birds and fore and hind limbs in mammals.

1.2 The specification of haltere shape and size by Ubx

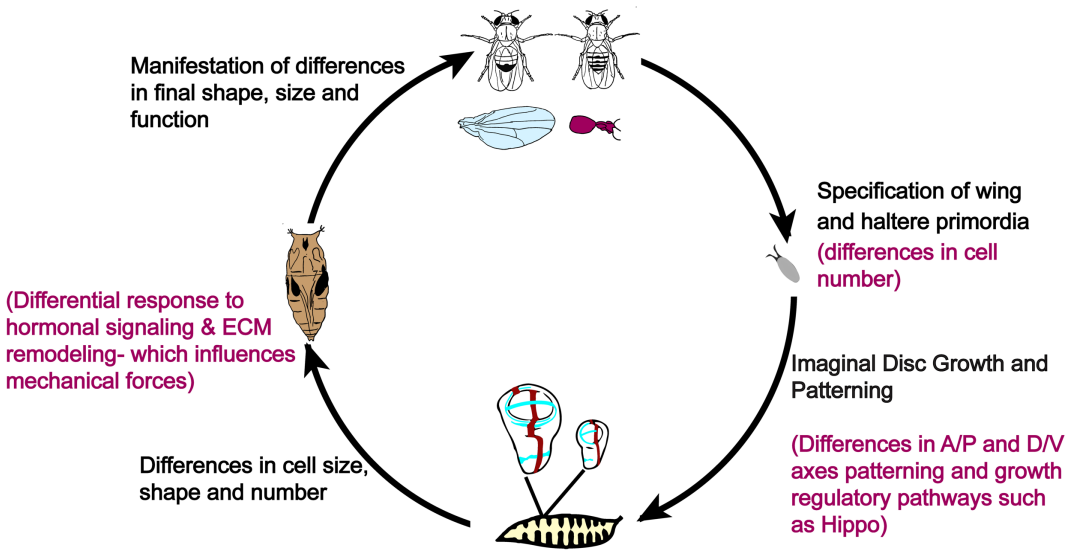
In *Drosophila*, Ubx specifies the haltere fate by modulating the expression of genes which are involved at multiple hierarchies of various wing development pathway (reviewed in Weatherbee et al., 1998). Understanding the events that control wing morphology, thus, is key to understanding the haltere specification program mediated by Ubx.

The wing in adult *Drosophila* is a two layered structure. The wing is flat and bears one trichome (wing hair) per cell. These trichomes are widely separated as wing cells are much larger and flatter. In contrast, the haltere is bulbous with the dorsal and ventral compartments not in physical contact with each other. Unlike wing cells, the haltere cells are more bulbous and contain multiple trichomes per cell (Roch and Akam, 2000). These differences in wing and haltere morphology which manifest in adult Drosophila are a result of differences in developmental events starting at the embryonic level.

The wing primordium, consisting of a group of cells, is specified at the embryonic level. These cells further organize themselves into a single sheet of epithelium. This structure which is termed as the wing imaginal disc, later gives rise to the dorsal and ventral wing blades, the

hinge region as well as the veins and sensory bristles of the wing margin. Two major patterning events along the antero-posterior (A/P) axis and dorso-ventral (D/V) axis control the wing development. The hinges and dorsal thorax in the adult body are specified by patterning events along the proximo-distal axis.

The haltere primordia, like the wing, is also specified at the embryonic level. However, the haltere primordia itself is made up of fewer cells. At this stage, while there is no difference between wing and haltere primordial cells in terms of shape, size and polarity, modulation of growth and patterning events occurring along the A/P and D/V axis eventually lead to differentiation of haltere discs into structurally as well as functionally different structures in comparison to wings (Fig 1.3). In the sections that follow, we provide a succinct account of the patterning events that occur in the wing imaginal disc and how Ubx modulates these corresponding events in the haltere imaginal discs.



Khan and Dilsha, 2020

Fig 1.3: Ubx control of events modulating the haltere fate in Drosophila

Figure showing the life cycle of Drosophila and key events involved in wing specification. Shown in magenta are the traits/events that are regulated by Ubx to specify the haltere. At the embryonic level, Ubx modulates differences in cell number. During larval stages, Ubx regulates AP and DV patterning events which impinge on cell size and shape change during the prepupal stages. At the pupal stage, Ubx modulates hormonal signaling and ECM remodeling to manifest differences in appendage shape and function.

1.2.1 Modulation of Antero-posterior patterning events by Ubx

In the wing imaginal discs, the cells expressing the Engrailed (En) protein define the posterior compartment. En induces the expression of the signaling molecule Hedgehog (Hh) which then diffuses into the anterior compartment. Hh further activates Patched (Ptc), Smoothened (Smo), Cubitus interruptus (Ci), Knotted (Kn) and Decapentaplegic (Dpp). Dpp acts as a long-range morphogen, activating several wing patterning genes like Spalt (Sal) and bifid (Omb). Cell proliferation in the wing disc is also dependent on proper positioning of cells receiving varied levels of Dpp (Aza-Blanc and Kornberg, 1999; Rogulja and Irvine, 2005; reviewed in Ruiz-Losada et al., 2018). Thus, the role of Dpp is central to wing development.

In the haltere imaginal discs, Ubx modulation of the A/P patterning events commences at the level of Dpp expression. Ubx directly represses *dpp* expression at the transcript level (Mohit et al., 2006). Ubx also upregulates the receptor of Dpp, Thick-veins (Tkv), in the A/P boundary. This leads to internalization of Dpp within the cells where it is made, thereby hindering its diffusion away from the A/P boundary. Additionally, Ubx also represses the expression of Dally, which acts as a facilitator for Dpp diffusion away from the A/P boundary(Crickmore and Mann, 2007, 2006; Makhijani et al., 2007). These modulation of Dpp expression and diffusion lead to its asymmetric distribution in the anterior and posterior compartment, thereby leading to asymmetry in cell proliferation between the two compartments. This is evident from the fact that in wing imaginal discs the anterior and posterior compartment are of equal size, however, in haltere imaginal discs the posterior compartment is much smaller (almost 3 times) as compared to the anterior compartment.

Ubx also regulates the expression of downstream components of the Dpp pathway like Dad, Spalt major and DSRF, thereby modulating the A/P patterning events at multiple levels in the haltere imaginal discs (Galant et al., 2002; Hersh et al., 2007; Mohit et al., 2006; Weatherbee et al., 1998).

1.2.2 Modulation of Dorso-ventral patterning events by Ubx

The proper patterning and growth of wing imaginal discs along the dorso-ventral (D/V) axis is determined by the D/V boundary which acts as an organizing center (Diaz-Benjumea and Cohen, 1993). The dorsal compartment of wing discs is specified by Apterous (Ap) gene, which further activates Serrate (Ser) and Fringe (Fng). Ser and Fng activate the Notch protein which responds to signals coming from the ventral compartment and specifies the D/V boundary (Brook et al., 1996; Ruiz-Losada et al., 2018). Notch also activates downstream

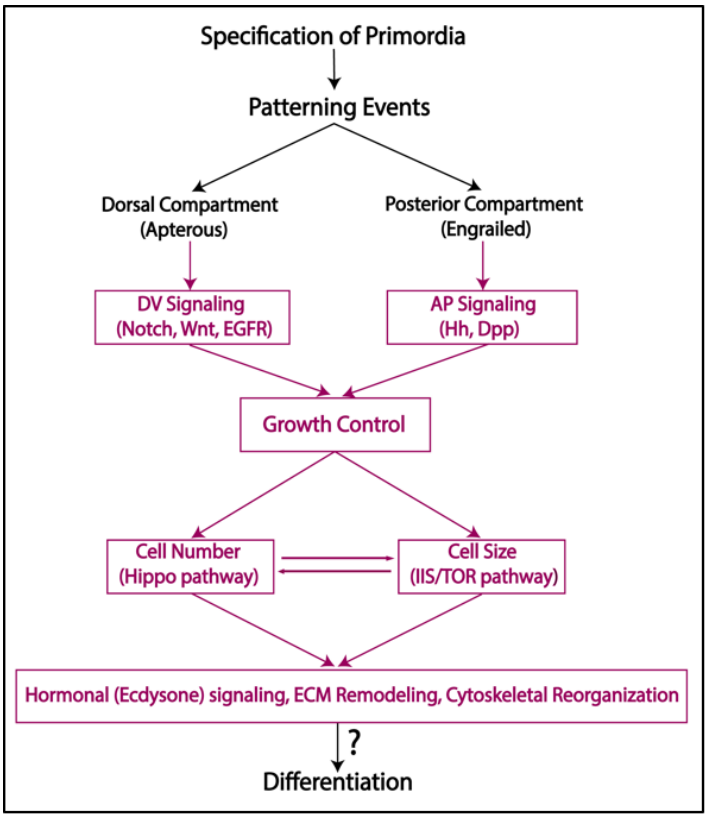
targets like Wingless (Wg), Cut (Ct) and the boundary enhancer of Vestigial (Kim et al., 1996). Wg further activates Vein (Vn), Achaete (Ac), Distal-less (Dll) and Vg in non-DV cells. The expression of Vg is very critical for the development of the wing. The *vg* gene being a pro-wing gene, any ectopic expression of this gene leads to ectopic wings. Thus, the expression of vg is tightly regulated in the wing disc.

In haltere discs, the expression of Ap, N, Ser and Vg (D/V boundary-specific) are unaltered. The expression of Wg, is repressed by Ubx, however, this is only in the posterior compartment. While it is still not clear how Ubx represses the expression of downstream targets of Wg in the anterior compartment, genetic experiments suggest that Ubx represses the wing identity in this compartment non-cell autonomously (Shashidhara et al., 1999). As previously stated, the *vg* gene is a pro-wing gene. Thus, the repression of Vg is critical for haltere specification. Since Vg is activated by N in the D/V boundary and in non-D/V cells by Wg, Ubx inhibits the nuclear localization of Armadillo (Arm) (required for Wg diffusion) in the anterior compartment, while also repressing events downstream of N. Additionally, Ubx also represses the non-D/V expression of Vg directly. Ubx, thus, modulates genes involved at multiple levels of the D/V patterning cascade to repress the wing identity and specify the haltere identity.

1.2.3 Modulation of cell number, cell size and ECM by Ubx

In the wing imaginal discs, the EGFR signal transduction pathway involving RAS/MAPK is known to control various aspects of organ size including cell proliferation, tissue growth as well as apoptosis (Pallavi and Shashidhara, 2003). It has been previously observed that overexpression of components of the EGFR pathway like EGFR and Vn, cause severe haltere to wing transformations. Thus, regulation of such components, which are directly targeted by Ubx (for example vein and pointed) is important for the haltere specification process. Another important pathway involved in the control of organ size is the Hippo/Yorkie pathway, which is considered to be the master regulator of crosstalk between different regulatory pathways and controls organ size by regulating cell-proliferation, cell growth and apoptosis (Irvine and Harvey, 2015). Additionally, growth is also dependent on environmental factors like nutritional status, which are sensed by the IIS/Akt pathways. In the haltere imaginal discs, Ubx is known to downregulate the function of both Hippo/Yorkie as well as IIS/Akt pathways to specify proper size and organ shape in the T3 segment (Singh et al., 2015). In addition, Ubx also modulates the expression of the Mmp1 gene in halteres, thereby inhibiting the degradation of

Viking and thus specifying the bulbous shape of the haltere (De Las Heras et al., 2018). Taken together, Ubx regulates a large number of genes which are involved in the mediating proper growth and shape of the wings, thereby specifying the haltere fate.



Khan and Dilsha, 2020

Fig 1.4: Ubx regulates multiple genes at various levels of hierarchies involved in wing specification pathways

Diagram showing the various events that are regulated by Ubx to specify the haltere. Ubx controls AP and DV patterning events by repressing key genes like *dpp* and *notch*. Ubx also directly regulates cell number and cell size by regulating key components of the Hippo and IIS/TOR pathways. In the pupal stages, Ubx regulates ECM remodeling and hormonal signaling to specify the bulbous shape of the halteres.

1.3 Mechanism of gene regulation by Ubx

It is quite evident that Ubx regulates a large number of genes at various hierarchies of the wing development pathway to specify the haltere fate in the T3 segment (Weatherbee et al., 1998). Ubx is involved, both, in upregulation as well as downregulation of target genes. However, being a homeodomain containing transcription factor, Ubx, like all Hox proteins, is known to bind to a motif "TTAATKR" containing a TAAT core site with very low specificity and affinity in-vitro. This is in sharp contrast to the highly specific regulation of target genes achieved by Ubx in *Drosophila* halteres and this apparent paradoxical behavior of Ubx, like all Hox genes, has intrigued biologists for various decades.

Various attempts made in the past two decades have, thus, tried to understand two major questions relating to Ubx mechanism of function: i) Being a homeodomain containing transcription factor which recognizes a frequently occurring TAAT core site in the genome, how does Ubx identify which sites to bind on in-vivo; ii) Since Ubx is involved in both upregulation and downregulation of target genes, what are the molecular cues that help it to recognize which targets to upregulate or downregulate.

Previous studies carried out using single gene approaches identified the role of homotypic clusters (multiple binding sites for a transcription factor in tandem) in regulation of certain Ubx targets like *spalt* (*sal*) (Galant et al., 2002). Analysis of other Ubx targets like Thickveins (tkv) and dally also revealed the presence of multiple Ubx binding sites (Makhijani et al., 2007). Additionally, unpublished ChIP-sequencing studies performed for the Ubx protein in Drosophila halteres also indicate that a large number of genes have more than one binding site for Ubx. This suggests a possible mechanism where Ubx may bind with low affinity to "TTAATKR" sequences and multiple such binding sites might contribute to stronger binding of Ubx, thereby leading to regulation of the target gene. Recent studies carried out with the *shavenbaby* (*svb*) gene in *Drosophila* embryos, further demonstrate the role of multiple low affinity binding sites and how they are critical for the spatio-temporal regulation of target genes by Ubx (Crocker et al., 2015).

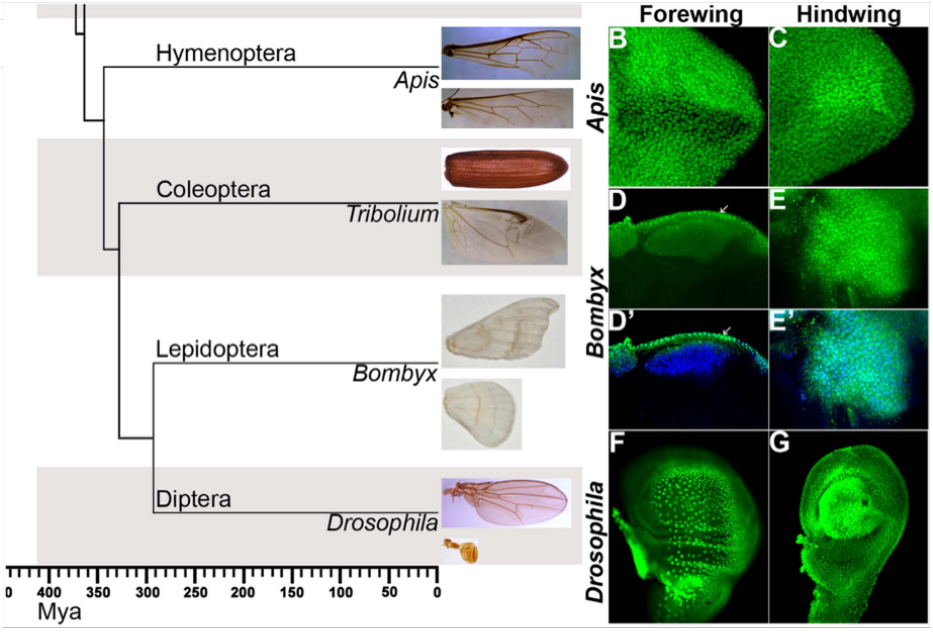
Association of Ubx with other transcription factors which can then stabilize its binding to the chromatin, is another way Hox protein like Ubx achieve specificity of target selection and regulation in Drosophila halteres. Previous studies using the Spalt major (*salm*) enhancer identified the role of association of Ubx and the MAD (Mothers against decapentaplegic) protein as critical for the repression of the salm enhancer in the halteres (C. M. Walsh and Carroll, 2007). Additionally, genome wide binding studies for the Ubx protein (ChIP-Chip)

carried out by Agrawal et al., 2011, identified that binding motifs for transcription factors like MAD, GAGA factor (GAF) and Adf1 are enriched in Ubx pulled down sequences. ChIP-qPCR studies further revealed that Ubx and GAF shared more than 100 target genes (Agrawal et al., 2011). Additionally, ChIP-western and EMSA studies revealed that Ubx and GAF interact at the protein level also, thus suggesting that association with transcription factors like GAF and MAD might be a possible mechanism by which Ubx can select and regulate its targets.

However, while several attempts have been made to understand the mechanisms of Ubx target selection and regulation, a comprehensive understanding of what guides Hox proteins like Ubx to select their targets specifically, in-vivo, still remains largely unanswered. More importantly, what remains elusive are the regulatory codes which determine whether binding of Ubx will upregulate a particular gene or downregulate it. Understanding these mechanisms will, thus, be critical for understanding morphological development by Hox proteins.

1.4 Ubx and Evolution of flight Appendages in different insect species

In Dipterans like *Drosophila melanogaster*, the Ubx protein is expressed, required and sufficient for haltere specification in the third thoracic (T3) segment. However, the presence of halteres in the dipteran lineage cannot be attributed merely to the expression of Ubx in the T3 segment. For example, Ubx is also expressed in the developing hindwings of the coleopteran, *Tribolium castaneum*, and is required for the development of hindwings in the T3 segment (Tomoyasu et al., 2005). In Lepidopterans, like *Precis coenia*, Ubx is expressed in T3 and regulates differences in the eyespot patterns and pigmentation between the fore and hindwings (Weatherbee et al., 1999). In Hymenopterans, like *Apis mellifera*, Ubx is expressed in both developing forewing and hindwing primordia, with the expression being slightly stronger in the developing hindwing (Prasad and Tarikere et al., 2016). Importantly, what is interesting is the fact that the DNA binding homeodomain region of Ubx is conserved across insect species, thereby adding another paradoxical level to the function of Hox genes; How do homologous Hox proteins, like Ubx, code for fundamentally different structures across different species?



Prasad and Tarikere et. al, 2016

Fig 1.5: Ubx expression in forewing and hindwing primordia of different species

Ubx is expressed in the third thoracic segment (T3) of *Drosophila melanogaster* and required for the specification of halteres. In *Tribolium castaneum*, Ubx is expressed and required for hindwing specification in T3. Ubx is also expressed in hindwing primordia of *Bombyx mori*. In *Apis mellifera*, Ubx is expressed in both forewing and hindwing primordia with the expression being slightly stronger in the hindwing primordia.

It is now clear from multiple studies that morphological differences coded by Hox proteins in different species do not arise due to major changes in target genes of the Hox protein, but rather in the regulation of common targets. Comparative analyses performed by Prasad and Tarikere et al., 2016, between the hindwing primordia of the Hymenopteran, Apis mellifera and developing halteres of the dipteran, Drosophila melanogaster, suggest that a large number of genes are common targets of Ubx between the two species. These common targets are genes which are involved in the wing morphogenesis pathway. A few of these genes like *vestigial* (*vg*), *Spalt* (*sal*) and *wingless* (*wg*), are differentially regulated by Ubx in *Apis* and *Drosophila*. While *vg, sal* and *wg* are expressed in both the forewing and hindwing primordia of *Apis*, Ubx represses the expression of these genes in the haltere imaginal discs in *Drosophila* to specify the halteres, thus bolstering the fact that mere binding of Ubx is not sufficient for the effective repression or activation of target genes. Careful inspection of the enhancer

sequences of the *vg* gene from *Apis* and *Drosophila* reveal a binding site for the Adf1 binding site which was present in the *Drosophila* version only but not in *Apis*. While experiment validation for this difference in Adf1 binding site between the two enhancers remain inconclusive, it is possible that changes in the enhancer regions of major wing patterning genes might have brought them under the negative regulation of Ubx in the dipteran lineage, thereby leading to haltere specification.

While other possible mechanisms like the role of different linker regions between Ubx from different species and dosage of the Ubx protein in hindwing primordia of different species also seem plausible, these need to be empirically tested in the context of morphological diversification during the course of evolution by Hox proteins.

Objectives of the Study

1. To identify in-vivo binding motifs for the Ubx protein in *Drosophila melanogaster* halteres using ChIP sequencing strategy
2. To elucidate the functional role of a strong binding motif for Ubx containing a TAAAT core in regulation of target genes in *Drosophila* halteres
3. To understand the role of the strong binding Ubx motif containing a TAAAT core in regulation of an enhancer of the *vestigial* gene from *Apis mellifera* in transgenic *Drosophila*
4. To compare mechanisms of target selection of the Ubx protein between *Drosophila* halteres and *Apis* hindwings
5. To understand the mechanisms governing recognition of activated vs repressed genes by Ubx

Chapter 2

Genome wide binding of Ubx in *Drosophila melanogaster* halteres

2.1 Introduction

The diversification of Hox gene function and regulation seems to be a critical factor for the evolution of body plans in the metazoan lineage. Hox genes encode for transcription factors containing a homeodomain region and are known to specify cell fates along the anterior-posterior axis of an organism (Lewis, 1978; McGinnis and Krumlauf, 1992). The role of Hox genes are critical to the development of an organism. This is best evident from the fact that any alterations in Hox gene function lead to severe homeotic transformations(Lewis, 1978; Struhl, 1981). Thus, changes in organ identity brought about by Hox gene mutations has served as a paradigm to understand mechanisms of morphological development.

Much of what we know about Hox gene regulation and function comes from studies carried out in *Drosophila melanogaster*. In *Drosophila* the second thoracic segment (T2) bears a pair of wings, whereas the third thoracic segment (T3) bears a pair of modified wing structures known as halteres. While the wing is still considered to be a Hox free state, the halteres are specified by the Hox protein, Ultrabithorax (Ubx) (Carroll et al., 1995). Loss of Ubx from developing halteres lead to haltere to wing transformations. On the other hand, overexpression of Ubx in developing wings lead to wing to haltere transformations (Lewis, 1978), suggesting that Ubx is necessary and sufficient for the haltere specification process. Multiple studies carried over the last few decades have identified that Ubx regulates a large number of downstream targets to specify the halteres. However, being a homeodomain containing protein, Ubx binds to a recognition motif containing a TAAT motif with low affinity and specificity. While it has been suggested that TALE (Three amino acid loop extension) class of transcription factors like Extradenticle (Exd) help Ubx bind with greater specificity (reviewed in Merabet and Galliot, 2015), certain appendages like the halteres in *Drosophila* do not require the input of TALE class of proteins. Thus, the mechanisms by which Ubx selects its targets are still not well understood.

Identifying all direct targets of Ubx as well as the enhancer sequences bound by Ubx in-vivo, are important to understand the mechanism of target selection. Previous studies to identify direct targets of Ubx have been limited to single gene or microarray-based approaches (Agrawal et al., 2011; Choo et al., 2011), which lack resolution. To circumvent this problem, we generated a high quality and high-density genome wide binding data for the Ubx protein in halteres using ChIP-sequencing. Our results suggest that, in-vivo, the Ubx protein binds to a motif (MATAAATCAY), containing a TAAAT core site (hereafter known as the TAAAT motif). We, however, did not see an enrichment of motifs containing a TAAT core site. We

performed extensive bioinformatics analyses to investigate whether the TAAAT motif serves as a recognition motif for Ubx to differentiate between upregulated and downregulated targets. However, we found the prevalence of this motif to be similar across upregulated and downregulated enhancers, suggesting that the TAAAT motif does not aid Ubx in discriminating between activated vs repressed targets . Taken together, our results uncover a novel binding motif which seems to be the preferred binding site for Ubx in *Drosophila* halteres.

2.2 Results

2.2.1 Chromatin Immunoprecipitation

Chromatin immunoprecipitation (ChIP) is an antibody-based technique used to identify chromatin binding sites of a particular transcription factor (reviewed in Das et al., 2004). In a ChIP assay, proteins are immobilised onto DNA using a crosslinking agent, followed by sonication to generate smaller DNA fragment and finally immunoprecipitation with protein-specific antibodies. ChIP assays can be coupled to various other techniques like PCR and qPCR to validate binding of a protein to known regions. Alternatively, coupling of ChIP reactions to massively parallel sequencing allows the identification of genome wide binding profiles for the protein to the chromatin at significantly higher resolutions as compared to ChIP-microarray approaches.

We performed ChIP for Ubx in third instar halteres of *Drosophila melanogaster* and coupled it to massively parallel sequencing. For each biological replicate, we collected 1000 haltere discs and 500 wing discs and generated three such independent biological replicates. The sonication process was standardised using mixed discs dissected from third instar larvae suggested that 25cycles at high frequency using the Covaris sonicator yielded optimum chromatin fragment length (Fig 2.1).

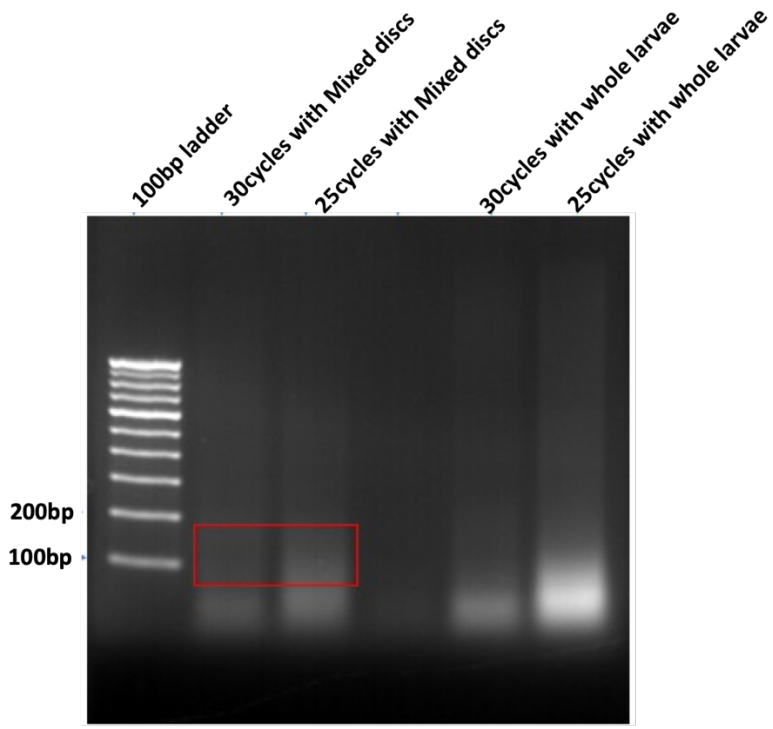


Figure 2.1: Standardization of sonication

The sonication was standardized on Covaris S2 using third imaginal discs isolated from third instar wandering larvae. Enrichment of DNA fragments around 150bp mark was observed after 25 high duty cycles.

Consequently, chromatin immunoprecipitation was performed using a polyclonal antibody raised against the N-terminal region of Ubx (Agrawal et al., 2011), for both control and test samples. ChIP PCR was performed for known targets of Ubx and revealed specific enrichment in only haltere discs and not in control or wing samples (Fig 2.2), thus, indicating the success of the experiment. Similar results were observed for all three biological replicates. ChIPed DNA was processed for Library Preparation and sequenced using the Hiseq 2500 V4 platform.

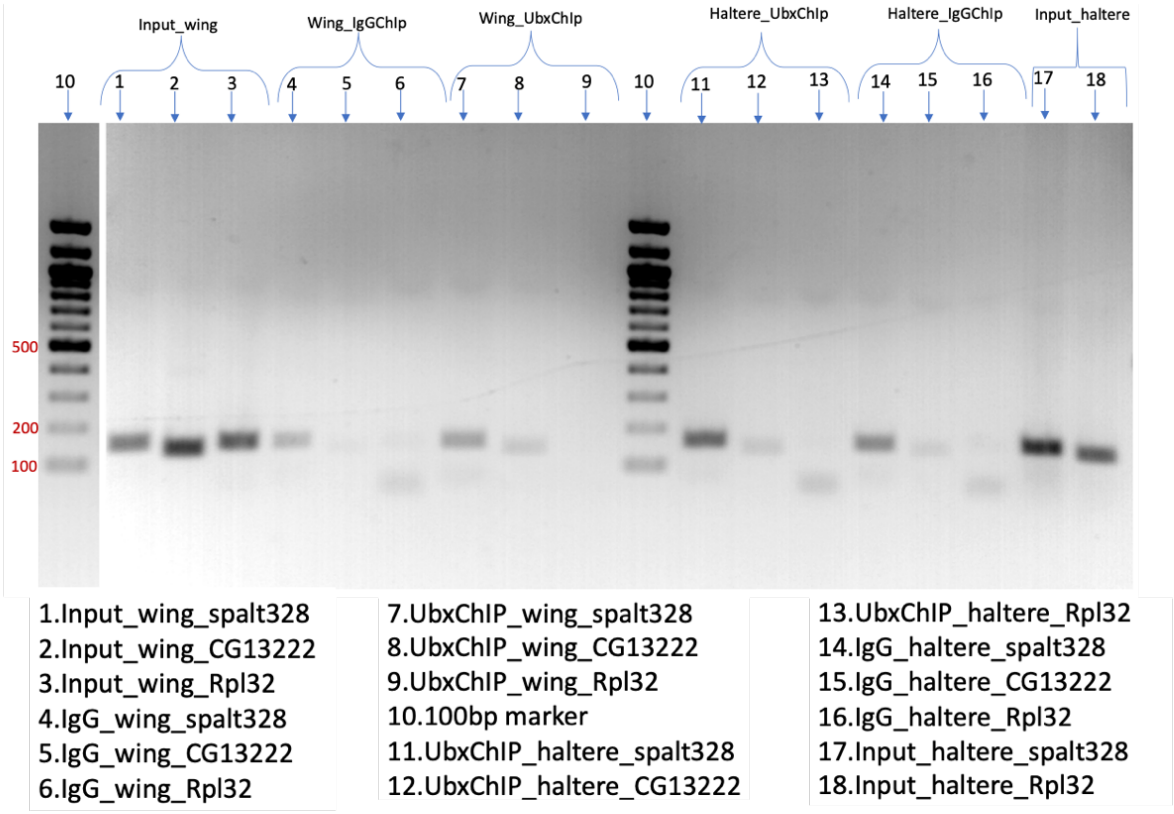


Figure 2.2 Validation of Chromatin immunoprecipitation using ChIP-PCR

ChIP-PCR for primers designed against known Ubx bound regions (hot regions) (sal328 element and edge enhancer of *CG13222*) as well as regions not bound by Ubx (cold regions) (enhancer of *Rpl32*). Genomic enrichment of hot regions, but not for cold regions, was observed in Ubx pulled down sequences (ChIP) as compared to IgG control. Enrichment was observed for all three sites in Input DNA

2.2.2 Identification of direct targets of Ubx in *Drosophila* halteres

Fastq files containing information of genome wide Ubx binding in the haltere and wing imaginal discs were obtained after sequencing of ChIPed DNA. We checked the quality of sequencing using FASTQC analyses and found that all sequencing reads passes quality control tests.

Next, raw FASTQ files were trimmed for adapter sequences using the Trimmomatic software and further aligned to the dm6 genome. Duplicate reads were filtered out and peak calling performed using MACS2 and PEPR software. To identify high confidence Ubx binding sites, we employed MACS2 with a cutoff criterion (FDR 0.05, maxdup 1) to call peaks in all

three independent biological replicates and chose only those peaks which were present in at least two replicates. Our analyses reveal 2171 high confidence binding sites for the Ubx protein, of which 838 peaks are found to be present in all three replicates. Annotating the peaks to their nearest TSS suggest that 1236 genes are direct targets of Ubx in the haltere imaginal discs. Among such targets were genes known to be regulated by Ubx like *salm, dally*, *np, CG13222, tkv, vg and sb*. Genomic tracks produced using the IGV software suggest occupancy of Ubx binding near the TSS of known targets (Fig 2.3A, Fig 2.3B), thus validating our dataset.

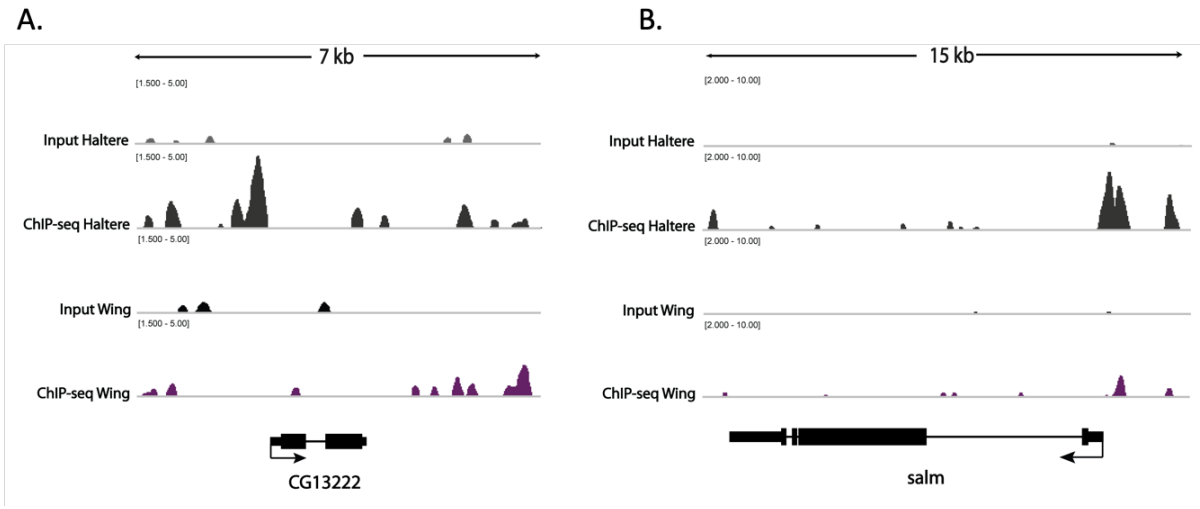


Figure 2.3. Occupancy of Ubx near TSS of known target genes

Ubx shows greater occupancy on chromatin upstream of TSS of known targets like *CG13222* (A) and *spalt* (B) in haltere imaginal discs as compared to input DNA. Ubx occupancy was not observed in chromatin from wing imaginal discs.

For a large number of targets, we observe multiple Ubx binding to regulatory regions. This was in accordance with previous studies which suggest that Ubx utilizes multiple monomeric binding sites for specificity of binding as well as target regulation (Galant et al., 2002). We further compared our ChIP-seq data with published ChIP microarray data from Choo et. al (2011) and Pavan et.al (2011) and found that 416 and 148 genes are shared targets between the studies respectively, with 116 genes being common targets in all three studies (Fig 2.4A). In terms of binding architecture, Ubx binding sites were seen to be enriched majorly in Intronic (36.85%) and Intergenic (39.11%) regions, reasserting the fact that Ubx acts primarily

at the enhancer level (Fig 2.4B). We performed gene ontology of Ubx targets and observed that a large proportion of genes were involved in the wing patterning processes (Fig 2.5). This further remonstrates that Ubx specifies the wing development pathway by modulating a large number of genes involved in the wing specification program.

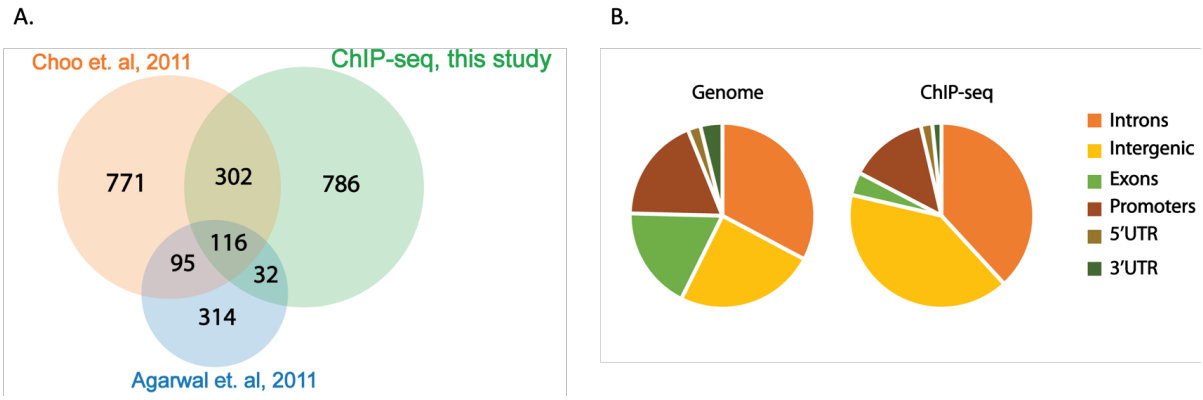


Fig 2.4: Annotation of Ubx bound regions and comparison of direct targets

A: Comparison of direct targets of Ubx identified in this study and published ChIP-microarray studies.

B: Ubx majorly binds at intronic and intergenic regions of the genome in *Drosophila* halteres

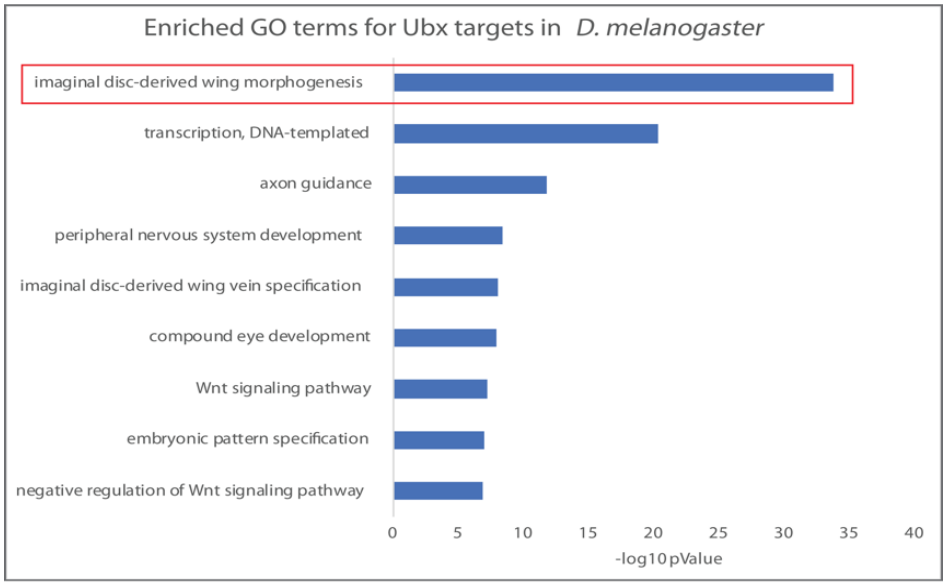


Fig 2.5: Gene ontology analyses of Ubx target genes
A large proportion of genes involved in wing patterning pathways are targets of Ubx in *Drosophila* halteres.

2.2.3 Identification of an in-vivo Ubx binding motif in *Drosophila* halteres

All Hox proteins, including Ubx, are known to bind, in-vitro, to a sequence containing a core TAAT motif. To identify what kind of recognition motif are bound by Ubx in *Drosophila* halteres, we performed de novo motif analyses on high confidence ChIP peaks. Our analyses reveal that a recognition motif (MATAAATCAY) with a TAAAT core sequence (hereafter referred to as TAAAT motif) is enriched by Ubx in *Drosophila* halteres (Fig 2.6A). We generated the position weight matrix (PWM) (refer to appendix1) for the TAAAT motif and calculated its frequency in Ubx pulled down sequences using the FIMO software from MEME suite. We also calculated the frequency of the TAAAT motif in the entire *Drosophila* genome (repeat masked) using the same approach. Our results reveal that the frequency of the TAAAT motif was 1.7 times greater in Ubx pulled down sequences as compared to the *Drosophila* genome (Fig 2.6B).

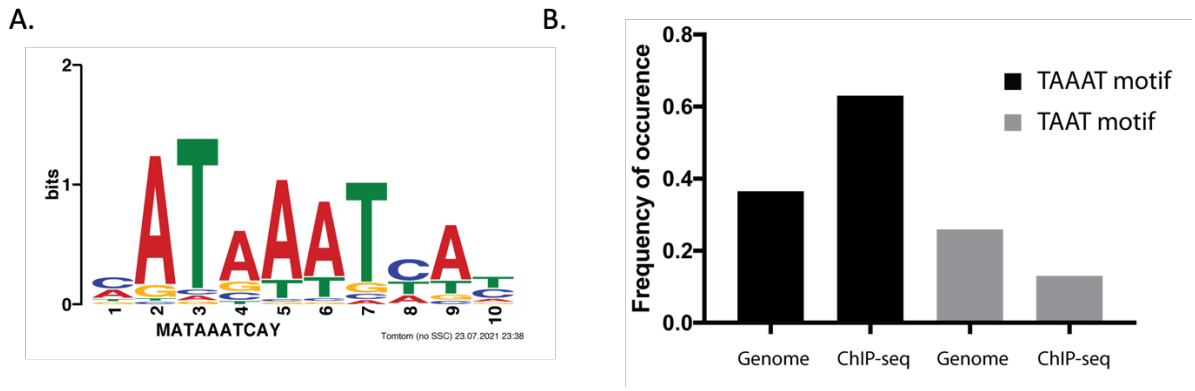


Fig 2.6: Motif analysis for Ubx pulled down sequences in *Drosophila* halteres

A: The DNA motif recognized by Ubx in *Drosophila* halteres

B: The frequency of the TAAAT motif is 1.7folds greater in Ubx pulled down sequences as compared to the whole *Drosophila* genome. The TAAT motif is not found to be enriched in Ubx pulled down sequences.

Interestingly, denovo analyses did not reveal enrichment of the recognition sequence containing a TAAT core sequence (hereafter referred to as TAAT motif), which is bound by Ubx in-vitro. Additionally, searching for the PWM of the TAAT motif (Noyes et al., 2008, Appendix 1), we found that the frequency of the TAAT motif was lower in Ubx pulled down sequences as compared to the entire *Drosophila* genome. This suggests that in the context of haltere specification, the TAAAT motif is more relevant than the TAAT motif for selection of targets by Ubx. We also observed enrichment of binding sites for transcription factors like Trl (Trithorax-like/GAGA factor), pho (pleiohomeotic) and brk (brinker) (Table 2.1), suggesting a possible role of these proteins as co-transcription factors of Ubx.

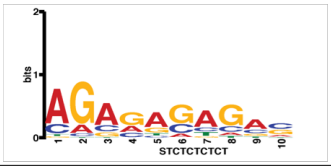
Enriched Motif in denovo motif analyses	Transcription factor Predicted to bind the motif
	Pho/Phol (Pleiohomeotic)
	GAGA factor
	Brk (Brinker)

Table 2.1. Transcription factors motif enriched in Ubx ChIP-seq
De-novo motif analyses of Ubx pulled down sequences in *Drosophila* halteres reveal enrichment of binding motifs for transcription factors like Pho, GAGA and Brk

2.2.4 Investigating the role of TAAAT motif in identifying activated vs repressed targets by Ubx

In *Drosophila*, Ubx specifies the haltere fate by regulating a large number of target genes. Previous studies have revealed that Ubx is able to both upregulate as well as downregulated target genes. However, the mechanisms which help Ubx recognize which targets to upregulate and which targets to downregulate are not well understood.

Since genome wide analyses reveal that a motif with a TAAAT core is enriched by Ubx in *Drosophila* halteres, we wanted to understand whether the TAAAT motif is used for identifying activated vs repressed targets by Ubx. To achieve this, we needed to examine the relative enrichment of the TAAAT motif in Ubx bound regions of target genes (from ChIP-seq data) that are upregulated or downregulated by Ubx. In this context, we identified genes that are differentially expressed between the wing and haltere imaginal discs using RNA sequencing (detailed in Chapter 6). We compared the list of direct targets with that of differentially expressed genes in the wing and haltere. This provided us with direct targets of Ubx that are upregulated and downregulated in the haltere. We also identified genes which are putative targets of Ubx but are not differentially expressed in the haltere. We traced back the

Ubx bound regions which were annotated to these genes and categorized such regions into upregulated, downregulated or not-differentially expressed enhancers.

Next, we calculated the frequency of the TAAAT motif using the FIMO software in each of the enhancer categories using a 200bp region centred around the ChIP peak. For normalization, we calculated the frequency of the TAAAT motif in large number of random sequences from the *Drosophila* genome (detailed in Materials and Methods). We find that the frequency of the TAAAT motif was similar in enhancers of all three categories; upregulated, downregulated and not-differentially expressed (Table 2.2). The frequency was also similar to that observed for Ubx pulled down sequences when compared to the entire *Drosophila* genome. This suggests that while the TAAAT motif seems to act as a motif for target selection in *Drosophila* halteres, it does not seem to have an obvious role in identifying activated and repressed targets by Ubx.

	Relative enrichment of TAAAT motif
Drosophila ChIP-seq	1.726
Putative Ubx Response elements of genes upregulated in haltere	1.905
Putative Ubx Response elements of genes downregulated in haltere	1.646
Putative Ubx Response elements of genes equally expressed in wing and haltere	1.807

Table 2.2: The TAAAT motif does not aid in identifying activated vs repressed targets
The frequency of the TAAAT motif was similar in Ubx bound regions of targets that are upregulated, downregulated or not-differentially expressed between the wing and haltere. Relative enrichment was calculated by normalizing to frequency of the TAAAT motif in random background sequences from the Drosophila genome.

2.3 Discussion

Hox genes play a pivotal role in diversification of body plans in the metazoan lineage. However, since all Hox genes bind, in-vitro, with low specificity to a motif containing TAAT core sequence, the mechanisms of target selection by Hox genes still remain largely unanswered. Identifying binding sequences for Hox genes, in-vivo, as well as the genes they directly target, thus, are key to understand the target selection process.

In *Drosophila,* the Hox protein Ubx specifies the haltere fate in the third thoracic segment by repression of wing specification pathways. While several attempts in the past have been made to identify direct targets of Ubx, all of them had focused on single gene approaches or microarray-based methods which lack depth and resolution. To this end we have performed chromatin immunoprecipitation (ChIP) for Ubx in haltere imaginal discs and coupled it to massively parallel sequencing (ChIP-seq). Generation of three independent biological replicates coupled with a highly specific polyclonal antibody (raised against the N-terminal of Ubx), make the dataset a valuable resource for the scientific community, one that was not available earlier.

Our studies provide interesting revelations into a novel DNA recognition motif containing a TAAAT core site (GSCCATAAATHA) which is enriched by Ubx in *Drosophila* halteres. The frequency of this motif was found to be 1.7folds greater in Ubx pulled down sequences as compared to the entire genome. This suggests that the TAAAT motif acts a preferred binding motif for Ubx in *Drosophila* halteres. While previous studies by Slattery et al., 2011 have indicated that the TAAAT motif is a high affinity binding motif for the Ubx-Exd complex, the development of haltere capitellum does not require the input of the Exd protein. Thus, the functional significance of Ubx binding the TAAAT motif in halteres, in the context of target selection, remains to be evaluated and will be addressed in subsequent chapters.

Using extensive bioinformatic analyses, we found that the frequency of the TAAAT motif was similar between Ubx bound regions of genes that are upregulated, downregulated as well as not-differentially expressed between the wing and haltere. Additionally, these frequencies were similar to the frequency of the TAAAT motif in all Ubx pulled down sequences. Our results, thus suggest, that the TAAAT motif is used only for selection of targets by Ubx and not for recognizing which targets to upregulate or downregulate. Perhaps, the TAAAT motif facilitates binding of Ubx to the chromatin but upregulation or downregulation of target genes depend on other transcription factors present in the vicinity of Ubx binding (addressed in Chapter 6).

Chapter 3

Evaluating the role of the TAAAT motif in Ubx mediated activation of *CG13222* in *Drosophila* halteres

3.1 Introduction

The specificity with which Hox proteins bind DNA is in sharp contrast to their specificity of function in-vivo, which is also termed as the Hox paradox. While TALE class of transcription factors like Extradenticle (Exd) and Homothorax (Hth) confer specificity by acting as cofactors for Ubx, various distal appendages like halteres appear not to require the function of Exd and Hth, suggesting a cofactor independent mechanism for identifying activated vs repressed targets. Using a genome wide approach, we identified a motif containing a TAAAT core site (referred to as the TAAAT motif) to be enriched in Ubx bound region in *Drosophila* halteres. The TAAAT motif, however, was found to be equally represented in putative Ubx enhancers of targets that are upregulated and downregulated in the haltere, suggesting that it is not required to differentiate between upregulated and downregulated targets by Ubx.

Previous studies in the field have identified that the TAAAT motif is recognized by the Ubx-Extradenticle (Exd) complex, however, the development of the haltere capitellum is independent of the input of Exd. Using well studied target genes as models, we characterized the role of the TAAAT motif in Ubx mediated activation/repression in *Drosophila* halteres. In this context, we identified TAAAT binding motif in the "*edge*" enhancer of the *CG13222* gene which is upregulated by Ubx in the haltere. Our results reveal that not only is the TAAAT motif important for Ubx binding to the *edge* enhancer of *CG13222*, it is also critical for its activation in Drosophila halteres. Additionally, a single point mutation changing the TAAT motif to TAAAT motif was sufficient to bring the enhancer under added regulation by Ubx. Taken together, our studies suggest a critical role of the TAAAT motif in the Ubx mediated upregulation of the *CG13222* enhancer.

3.2 Results

3.2.1 The TAAAT motif is required for Ubx binding to the CG13222 enhancer

To investigate the functional role of the TAAAT motif, we first sought to evaluate the relative affinity and specificity of Ubx binding to this motif as against the TAAT motif. We chose two well-characterized targets of Ubx, Spalt (*sal*) and *CG13222*. Previous studies have reported that the enhancer of the *sal* gene has multiple TAAT motifs and that binding of Ubx to these motifs is responsible for the effective repression of the sal gene in halteres (Galant et al., 2002). The *CG13222* gene, on the other hand, is activated by Ubx in the haltere imaginal discs and does not show any expression in the wing imaginal discs. Previous studies have identified a

459 bp "*edge*" enhancer for the *CG13222* gene which contains two Ubx binding sites; site1 and site2 (Fig 3.1) (Hersh et al., 2007). Mutations within the site1 (hereby known as M1) led to a significant loss of Ubx binding *in-vitro* as well as its expression in the halteres whereas mutation within site2 (hereby known as M2) does not show a significant effect (Fig 3.1) (Hersh et al., 2007).

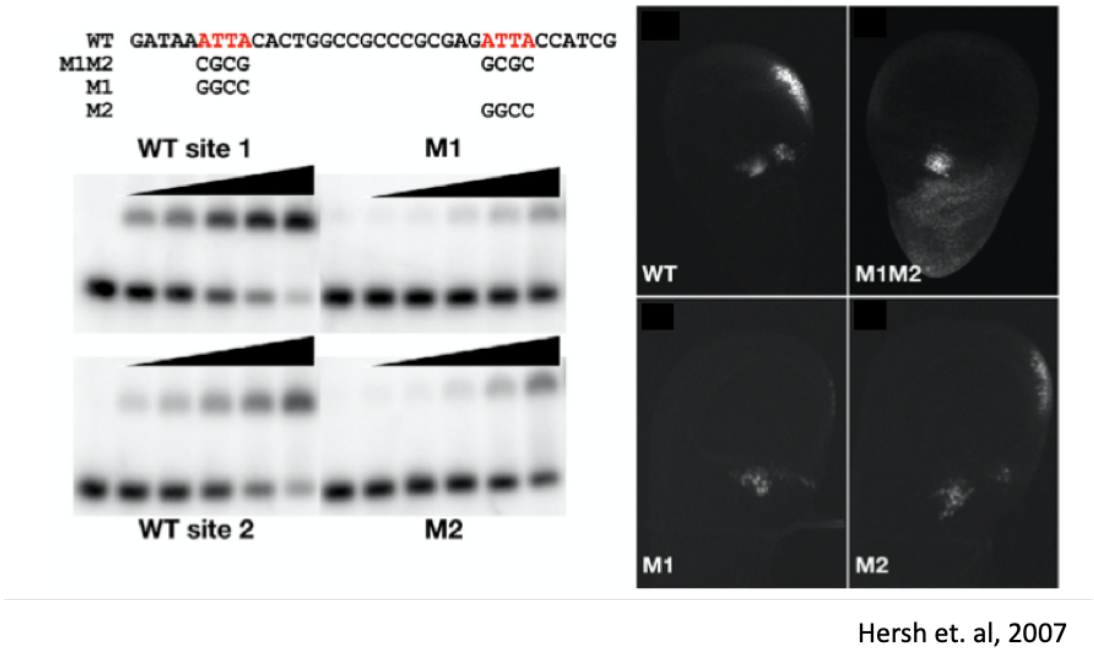


Fig 3.1: Ubx binding sites in the *CG13222* enhancer
Part of the *edge* enhancer of the *CG13222* gene containing the TAAT motifs (Hersh et. al, 2007). The site1 is critical for Ubx binding as well as activation of the enhancer in halteres whereas the site2 is dispensable.

Interestingly, we found that a TAAAT motif overlaps with the TAAT core in site1 (Fig 3.2) and the M1 mutant designed by Hersh et. al 2007, mutates both the sites (Fig 3.2). We designed mutations to specifically mutate the TAAAT core motif (M1_A) or the TAAT core motif (M1_B).

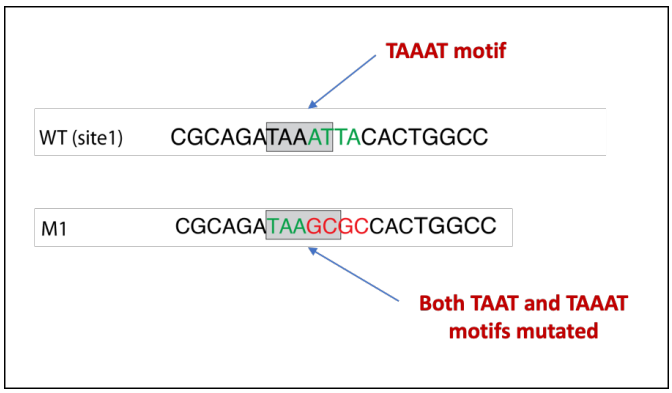


Fig 3.2: The *CG13222* enhancer has a TAAAT motif
The site1 of the *edge* enhancer of *CG13222* has both TAAT and TAAAT motifs. The M1 mutant generated by Hersh et. al, 2007 changed both the TAAT and TAAAT motifs.

We used a 29bp long probe for *sal* and a 21bp long probe covering site 1 of the *CG13222* enhancer for EMSA (Fig. 3.3A). We tested two different Hox proteins, Ubx and Sex combs reduced (Scr) on these probes for assessing DNA-binding specificity *in vitro*. Our results showed that both the Hox proteins bind on the probe derived from the *sal* enhancer, which contains only TAAT motifs, although with less affinity for Scr (probe Spalt 5/6 in Fig. 2B). This suggests that the TAAT motif is not bound exclusively by Ubx but also by other Hox proteins like Scr. In contrast, only Ubx could bind on the probe derived from the *CG13222* enhancer, which had a TAAAT motif, suggesting a higher level of DNA-binding selectivity when compared to the *sal* enhancer (probe WT (site1) in Fig. 3.3).

Next, we compared the binding of Ubx to the TAAT vs TAAAT motif which overlap in the CG13222 enhancer. We find that the mutation in TAAAT motif (M1_A), where the TAAT motif is intact, in site1 of *CG13222* led to a significant loss of Ubx binding. On the other hand, mutation in the TAAT motif (M1_B), where the TAAAT motif is intact, did not show significant reduction of Ubx binding, thereby suggesting that the TAAAT motif is bound with higher affinity as compared to the TAAT motif in the *CG13222* enhancer (Fig. 3.3B).

A.

Spalt 5/6	GAATCATATTAAGACGGGCACATTATAAA
WT (site1)	CGCAGATAAATTACACTGGCC
M1_A	CGCAGAGGGATTACACTGGCC
M1_B	CGCAGATAAATGGCACTGGCC

B.

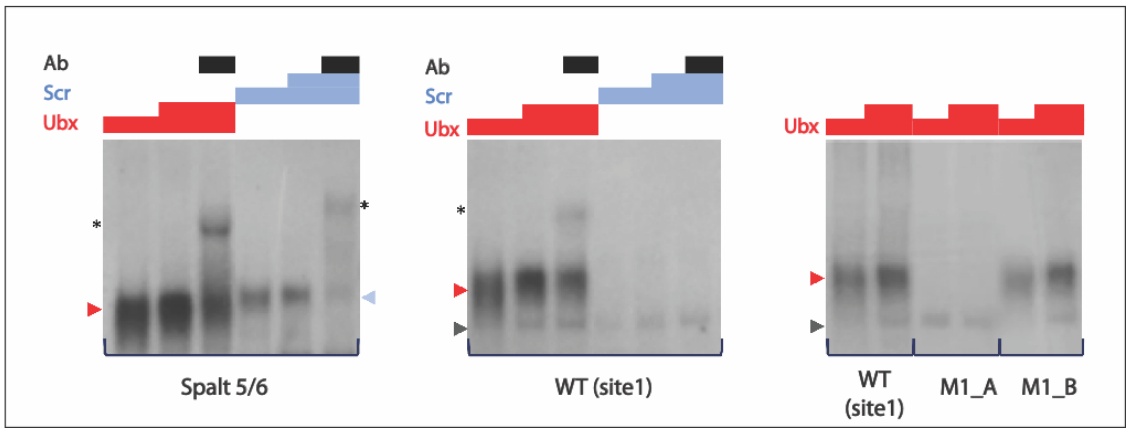


Fig 3.3: EMSA assays to evaluate the role of TAAAT motif in conferring affinity and specificity of binding to Ubx

A: Sequences of Spalt 5/6 and CG13222 probes used for EMSA. The TAAAT motif is specifically mutated in M1_A, whereas, in M1_B, the TAAAT motif is intact but the TAAT motif is mutated

B: While both Ubx and Scr bind to TAAT motif of *sal*, only Ubx binds to TAAAT motif of *CG13222* (WT site1). Binding of Ubx to WT Site 1 is severely reduced when the TAAAT motif is mutated to TAAT motif (M1_A), whereas no such effect is seen on mutating the TAAT motif (M1_B)

Next, we wanted to evaluate relevance of the TAAAT motif as compared to the TAAT motif on gene expression. To this end we employed Luciferase assays. In this context, we cloned the *edge* enhancer of *CG13222* enhancer and its various mutants into the pGL3-DE5 vector containing five dorsal binding sites upstream of the Luciferase reporter gene. In S2 cells, the WT *CG13222* enhancer is activated by a factor of 3 folds on Ubx induction (Fig 3.4).

Interestingly, the M1_A mutant, where only the TAAAT motif is mutated, shows a significant decrease of reporter activity (Fig 3.5), suggesting that the TAAAT motif is critical, not only for binding, but also for Ubx mediated activation of the enhancer in S2 cells.

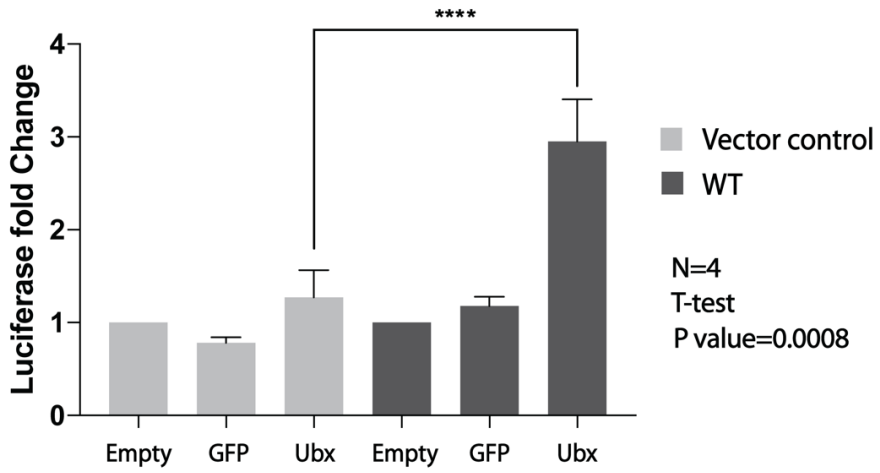


Fig 3.4: The *CG13222* enhancer is upregulated by Ubx in S2 cells

Luciferase assays indicate that the edge enhancer of CG13222 (WT) is activated in S2 cells on Ubx induction. Statistical test was performed using t-test (two tailed). Empty pGL3 vector was used as the vector control.

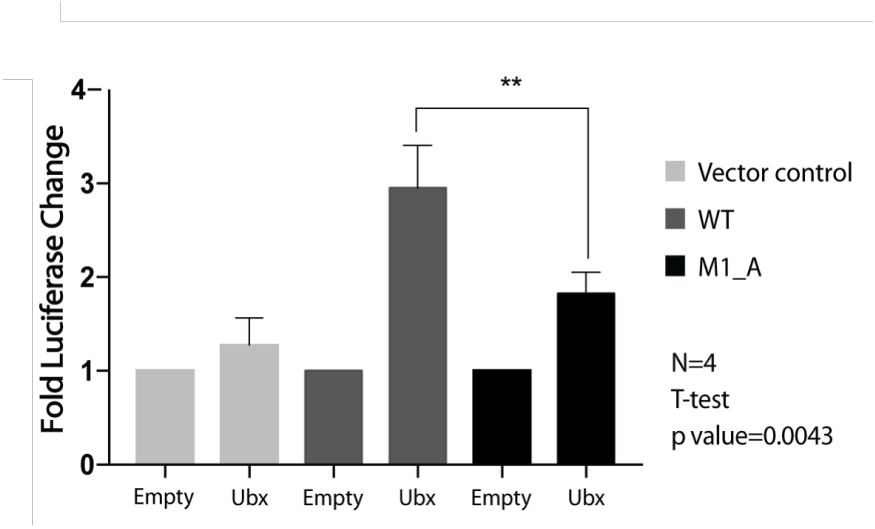


Fig 3.5: The TAAAT motif is critical for activation of *CG13222* in S2 cells

Mutation of the TAAAT motif in the *edge* enhancer of *CG13222* (M1_A) lead to significant reduction of Luciferase reporter activity in S2 cells as compared to the wild type (WT) enhancer. Statistical test was performed using t-test (two tailed).

2.2.2 The TAAAT motif is critical for Ubx mediated activation of *CG13222*

To examine if the TAAAT motif mediated activation of *CG13222* by Ubx also holds true in the haltere imaginal discs, we generated transgenic Drosophila strains carrying the WT and mutant enhancer constructs (M1, M1_A) upstream of a GFP reporter. 3rd instar larvae were cut, inverted and immunostained using anti-GFP antibodies and fluorescence intensity measurements performed using a confocal microscope. Consistent with previous studies, we found that the WT *CG13222* enhancer drives GFP expression along the posterior edge of the haltere imaginal discs but not in the wing imaginal discs (Fig 3.6).

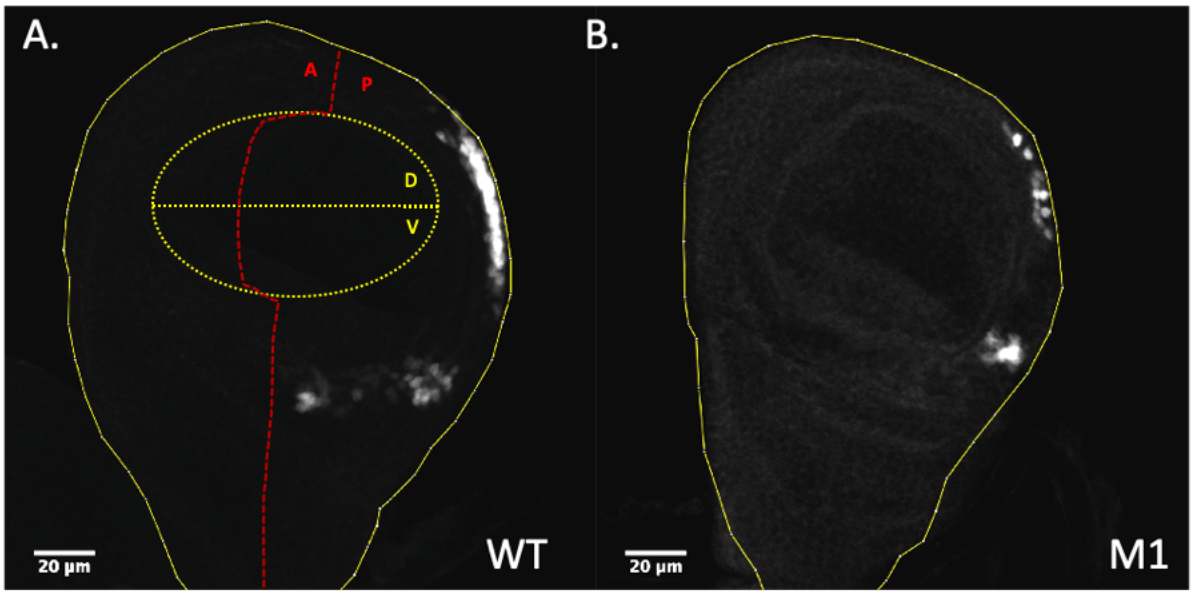


Fig 3.6: GFP reporter expression driven by *edge* enhancer of *CG13222*
The *edge* enhancer of *CG13222* drives reporter expression in the posterior edge of the haltere (A). The M1 mutant where both the TAAT and TAAAT motifs are mutated shows a loss of reporter expression (B)

Interestingly, the M1_A mutant, where the TAAAT motif is mutated and TAAT motif is intact, showed a marked reduction in GFP expression (Fig 3.7B). The phenotype for the M1_A mutant closely resembled that shown by the M1 mutant which has both the TAAT and TAAAT motifs mutated (Fig 3.7B). Quantification of mean fluorescence intensity revealed that the decrease in enhancer expression was significant (Fig 3.7C). Our results, thus, suggest that the TAAAT motif, and not TAAT motif, is critical for Ubx mediated activation of the enhancer in the haltere imaginal discs.

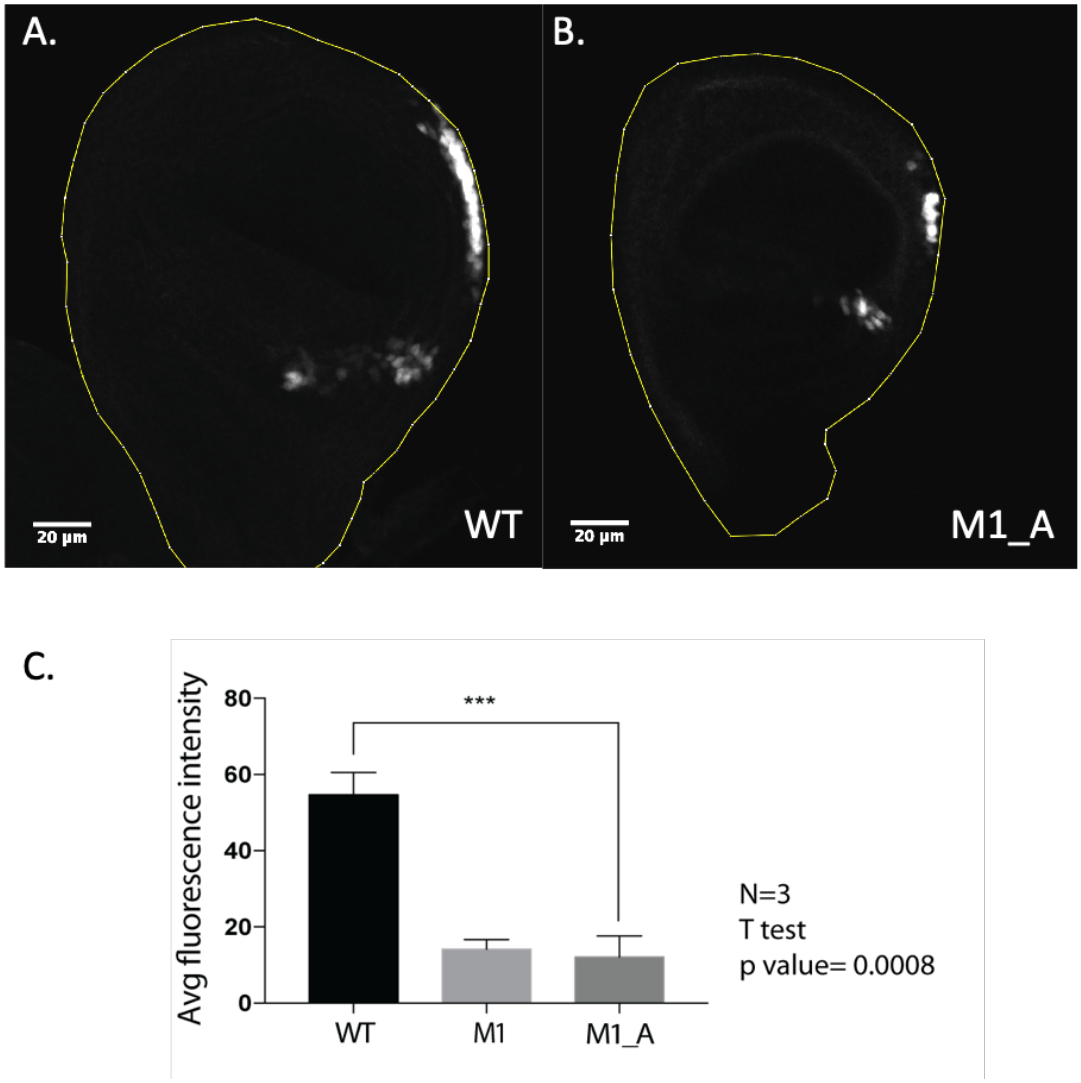


Fig 3.7: Role of TAAAT motif in Ubx mediated activation of *CG13222* in halteres
The *edge* enhancer of *CG13222* drives reporter expression in the posterior edge of the haltere (A). The M1_A mutant where both only the TAAAT motif is mutated shows a significant reduction in average fluorescence intensity of the GFP reporter (B, C). Statistical test was performed using t-test (two tailed).

2.2.3 The TAAAT motif brings the otherwise unresponsive site2 of the enhancer of *CG13222* under Ubx control

Earlier studies have demonstrated that mutations of the TAAT motif at site2 of the *edge* enhancer of *CG13222* does not show significant differences in Ubx mediated activation of the gene in *Drosophila* halteres (Fig 3.1). To get a better understanding of how TAAAT motifs may influence regulation by Ubx, we performed functional assays with mutant constructs

where we replaced the TAAT core on site2 with TAAAT (M2_A). In S2 cells, the M2_A construct drove a significantly reduced level of Luciferase activity as compared to the WT enhancer (Fig 3.8). Conversely, in haltere imaginal discs of transgenic *Drosophila* strain carrying the M2_A construct, we not only observed an enhancement of reporter expression but also found GFP expression in novel domains at the posterior edge of the haltere (Fig 3.9). This suggests that presence of TAAAT brought the site2 under Ubx regulation both *in-vitro* as well as in haltere imaginal discs. However, the difference in regulation of the mutant enhancer between S2 cells and halteres might be due to the difference between chromatin architecture between S2 cells and halteres.

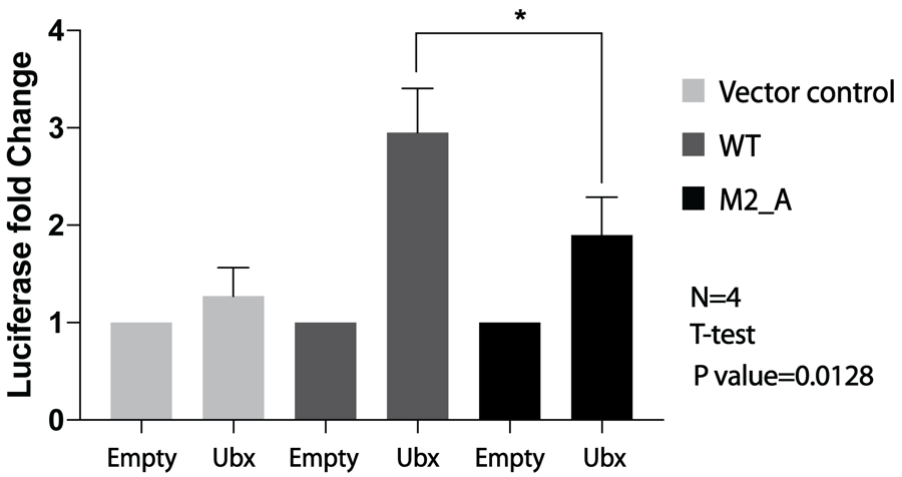


Fig 3.8: Role of the TAAAT motif in site2 of the *CG13222* enhancer in S2 cells
Mutation of the TAAT motif to TAAAT motif in the otherwise unresponsive site2 of *edge* enhancer of *CG13222* (M2_A) lead to significant reduction of Luciferase reporter activity in S2 cells as compared to the wild type (WT) enhancer. Statistical test was performed using t-test (two tailed).

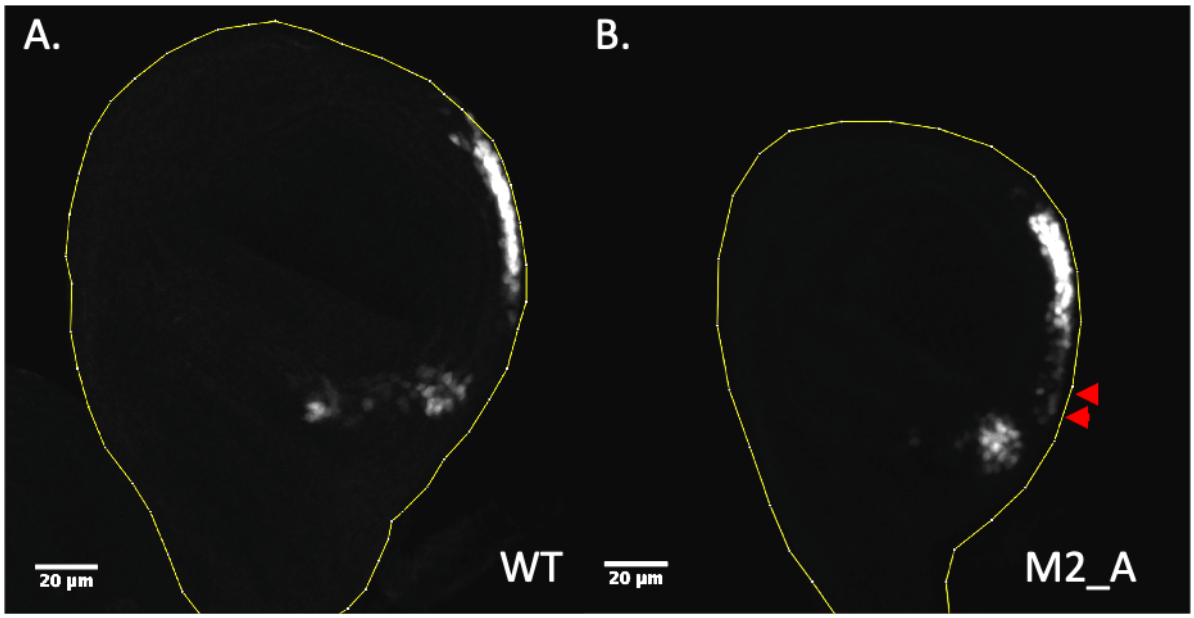


Fig 3.9: The TAAAT motif brings the site2 of the edge enhancer under Ubx control
The edge enhancer of CG13222 drives reporter expression in the posterior edge of the haltere (A). The M2_A mutant where the TAAT motif in site 2 is replaced by TAAAT (thereby leading to two TAAAT motifs) not only shows a significant increase in GFP reporter expression but also expression in ectopic regions (B) (shown by red arrows).

3.3 Discussion

Identifying mechanisms of target gene selection and regulation by Hox genes is key to understanding morphological evolution. While association with cofactors like the TALE group of transcription factors have been implicated in providing specificity of function to Hox proteins, certain distal appendages like halteres in *Drosophila* do not require the input of these co-transcription factors. Additionally, while low affinity binding sites are known to be critical for specificity of function for Hox proteins like Ubx, the role of a high affinity motif in this context is not well explored.

In this study, we have identified a motif containing a TAAAT core which is enriched by Ubx in *Drosophila* halteres and seems to be the preferred binding motif for Ubx in-vivo. The TAAAT motif was bound with higher affinity and specificity by Ubx as compared to the TAAT motif, as revealed through in-vitro EMSA assays. We also observe that this high affinity TAAAT motif is critical for the activation of the edge enhancer of the CG13222 gene in both *Drosophila* halteres as well as S2 cells. Additionally, the TAAAT motif was sufficient to bring an otherwise unresponsive binding site of Ubx (having a TAAT motif) under Ubx regulation.

The TAAAT motif, thus, is not only able to confer increased specificity of binding, but can also facilitate target gene selection which may be possible due to a more favorable chromatin structure for Ubx binding attained by the TAAAT motif as compared to the TAAT motif. Taken together, our studies reveal a high affinity binding motif for the Ubx protein in *Drosophila* halteres which is functionally relevant for the activation of a target gene during haltere development and can throw some light into mechanisms of target selection by Hox proteins.

Chapter 4

Functional significance of TAAAT motif in repression of enhancers of the *vestigial* gene from *Drosophila melanogaster* and *Apis mellifera*

4.1 Introduction

We observed that the TAAAT motif regulates the expression of *CG13222* in *Drosophila* halteres. Not only was the TAAAT motif critical for Ubx binding to the edge enhancer of CG13222 in-vitro, mutation of the TAAAT motif (and not the TAAT motif) led to a significant drop of expression driven by CG13222 enhancer in both S2 cells and haltere imaginal discs. This suggests that the TAAAT motif is functionally more critical than the TAAT motif in the context of upregulation of *CG13222* by Ubx.

We had previously shown that the frequency of the TAAAT motif is similar between enhancers of both upregulated and downregulated targets, indicating that the TAAAT motif is the favored binding site for Ubx in-vivo and might be used for target selection, but not for identifying activated and repressed targets. Therefore, it is probable that in addition to mediating Ubx mediated upregulation of targets (like *CG13222*), the TAAAT motif might be involved in Ubx mediated repression of targets. This was particularly important in terms of understanding developmental processes by Hox genes, as the specification of the haltere fate in *Drosophila* is mediated by repression of major wing patterning genes, like *vestigial*, by Ubx.

Additionally, wing patterning genes like *vestigial (vg)*, *spalt (sal)* and *wingless (wg)*, which are common targets of Ubx in *Apis mellifer*a and *Drosophila melanogaster*, are differentially regulated in the two lineages (Prasad and Tarikere et al., 2016). For instance, in *Drosophila*, *vg* is expressed and required for wing development and Ubx downregulates the quadrant enhancer of the *vg* gene (*quad-vg* enhancer) in halteres. In contrast, in *Apis*, *vg* is expressed in both forewing and hindwing primordia. Furthermore, an enhancer of *vg* identified in *Apis*, which was equivalent to the *quad-vg* enhancer in Drosophila, drove strong reporter expression in both wing and haltere imaginal discs in transgenic *Drosophila* and was not sensitive to Ubx regulation (Prasad and Tarikere et al., 2016). Since *vg* is a pro-wing gene whose repression is critical for the specification of haltere fate in *Drosophila*, we sought to understand the role of the TAAAT motif, (if any), in its Ubx mediated regulation in the halteres using the *quad-vg* enhancer. Additionally, understanding how Ubx differentially regulates the expression of the orthologous enhancers of *vg* in *Apis* and *Drosophila* can provide critical insights into mechanisms of morphological evolution.

We first scanned the quadrant enhancer of the *vg* (hereafter referred to as *quad-vg* enhancer) for the presence of Ubx binding motifs. We find that this enhancer has a TAAAT as well as a TAAT motif. We also scanned the enhancer of the *vg* from *Apis* (hereafter referred to as *Apis-vg* enhancer) and found that it contains only one Ubx binding motif, the TAAT motif.

Our results reveal that most mutations in Ubx binding motifs of the *quad-vg* enhancer led to a complete loss of reporter GFP expression from wing and haltere imaginal discs in transgenic *Drosophila*, suggesting that such sites might be relevant for the expression of the enhancer in the wing imaginal discs, and by a similar logic, for repression by Ubx in the haltere imaginal discs. More importantly, our studies reveal that changing the TAAT motif to TAAAT motif in the *Apis-vg* enhancer, was sufficient for the repression of the mutant enhancer specifically in the haltere pouch while not affecting its expression in the wing pouch. Taken together, our studies suggest the role of a microevolutionary change in the Ubx binding motif (TAAT to TAAAT) in the enhancer of the *vestigial* gene, which helped bring it under the control of Ubx in dipteran lineage and thereby facilitating the haltere specification process.

4.2 Results

4.2.1 Enhancers of *spalt* and *knot* do not have a TAAAT motif

We sought to investigate the role of the TAAAT motif in Ubx mediated downregulation of target genes. To this end, we screened enhancers of known Ubx targets for the TAAAT motif. The enhancers of Spalt (*sal*) and *knot* genes contain multiple TAAT motifs as indicated by previous studies. These genes are repressed by Ubx in the halteres (Galant et al., 2002; Hersh and Carroll, 2005). However, we did not find any TAAAT motifs within and/or near TAAT motifs. Intriguingly, inspection of Ubx occupancy in the haltere indicated a TAAAT motif present just upstream of the *spalt* TSS, however, further studies would be required to assess its importance in the context of *sal* regulation by Ubx.

4.2.2 The quadrant enhancer of the *vestigial* gene in *Drosophila* has a TAAAT motif

The *vestigial* (*vg*) gene is a pro wing gene whose ectopic expression of this gene leads to formation of ectopic wings. In haltere imaginal discs, Ubx represses the expression of *vg* by binding to its quadrant enhancer (*quad-vg* enhancer) and thereby repressing wing formation. However, how Ubx regulates the *quad-vg* enhancer is not known. To investigate whether the TAAAT motif has a role in the Ubx mediated repression of *vg* in the halteres, we performed motif scanning of the *quad-vg* enhancer using FIMO software. We identified a 25bp cassette containing two Ubx binding motifs; one TAAAT and one TAAT (Fig 4.1 A). Sequence alignment of the cassette containing the Ubx binding motifs revealed high conservation of both

TAAT and TAAAT among most *Drosophila* species (Fig 4.1 B), suggesting that the region might be important for the regulation of the *vg* gene.

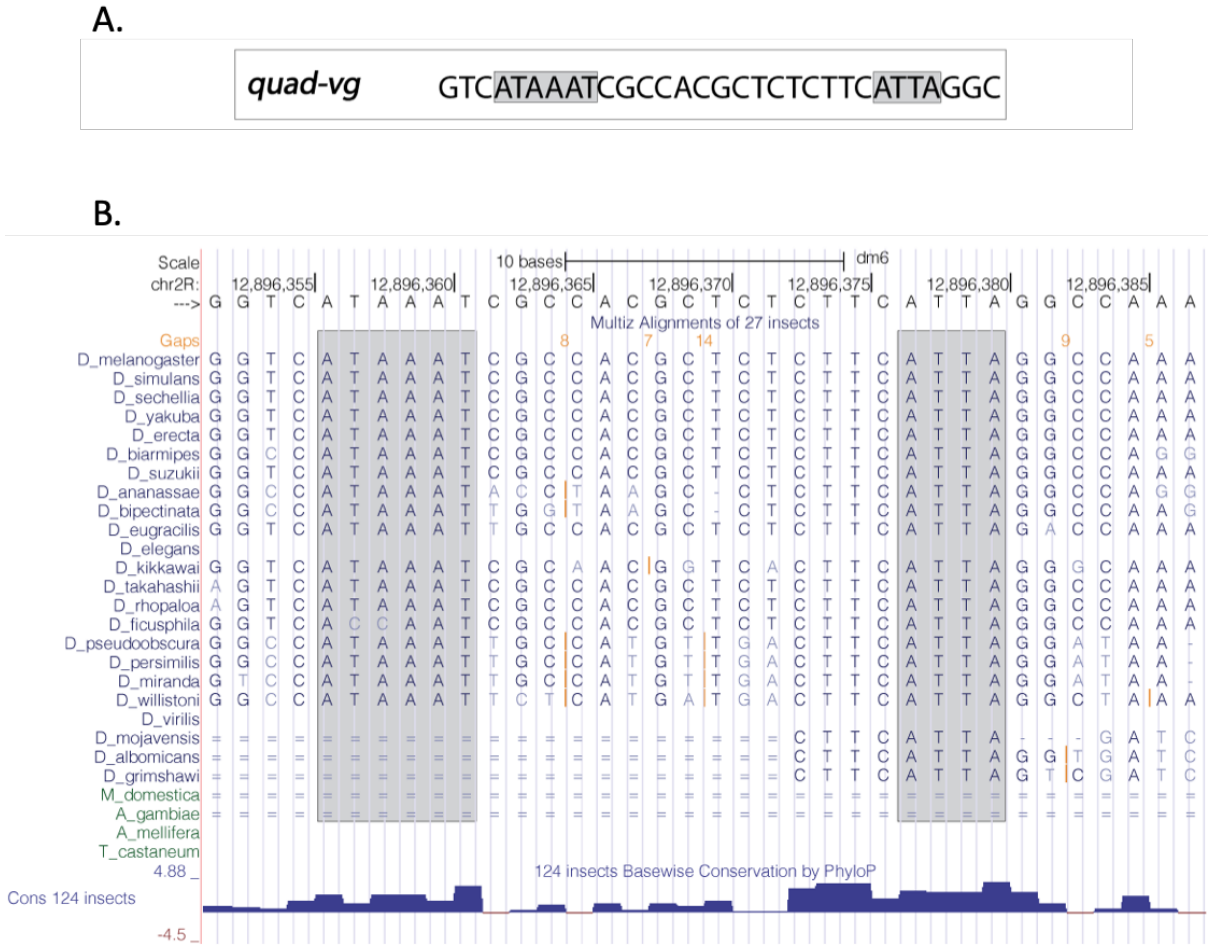


Fig 4.1: Ubx binding motifs in the quadrant enhancer of *vg (quad-vg)* from *Drosophila*

A: Part of the *quad-vg* enhancer gene containing the TAAAT and TAAT motifs identified using motif scanning by the FIMO software.

B: Sequence alignment of the cassette containing the TAAAT and TAAT motif of the *quad-vg* enhancer shows high conservation across most *Drosophila* species

To assess the functional role of the TAAAT motif of Ubx in repression of *quad-vg*, we designed several mutations (Fig 4.2 A) which alter either the TAAAT motif (*quad-vg_M*, *quad-vg_M1*) or the TAAT motif (*quad-vg_M2*) and cloned them upstream of a GFP reporter in the pH-Stinger-attb vector. We generated *Drosophila* transgenics carrying the wild type and

mutant *quad-vg* enhancers and performed immunostaining for GFP in third instar wing and haltere imaginal discs. Consistent with previous studies, we observed that the wild type *quad-vg* enhancer drives GFP reporter expression in wing imaginal discs, albeit at low levels (compared to earlier reported *quad-vg* lacZ reporter), but not in the haltere imaginal discs. However, any mutation in the quad-vg enhancer seems to render it inactive as observed from the near complete loss of reporter GFP expression in both wing and haltere discs for most mutant *quad-vg* constructs *in vivo* (Fig 4.2 B). While this could suggest that the any alterations in Ubx binding motifs of the enhancer lead to complete loss of activity of the enhancer element, further experiments need to be performed to back this observation

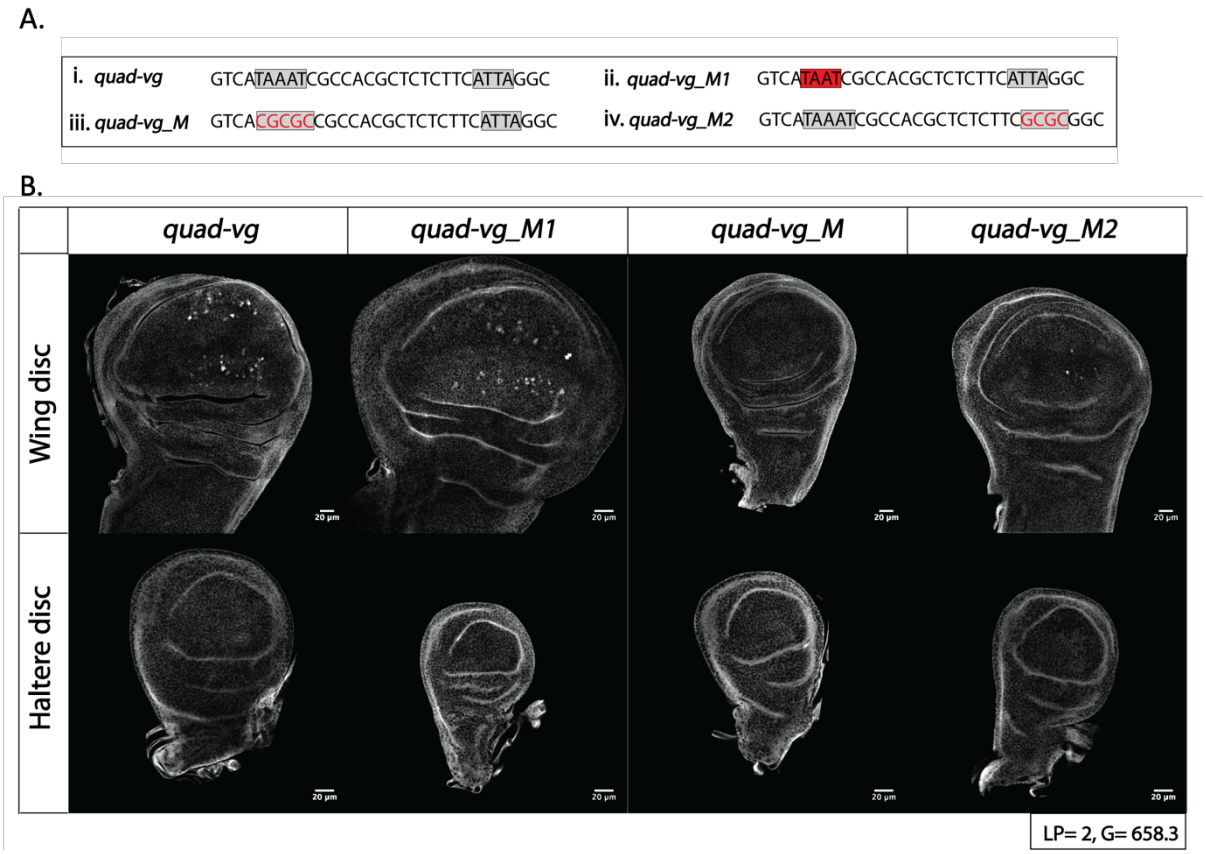


Fig 4.2: Role of TAAAT motif in Ubx mediated repression of *quad-vg* enhancer

A: Part of the *quad-vg* enhancer containing the TAAT and TAAAT motifs (i). Various mutations were designed to disrupt the TAAAT and TAAT motifs.

B: The *quad-vg* enhancer drives expression of the reporter GFP in the wing imaginal disc but not in the haltere imaginal disc. Most mutations in the TAAAT or TAAT motifs seem to render the enhancer inactive as seen from a complete loss of reporter gene expression in both wing and haltere imaginal discs. The laser power (LP) and gain (G) used for imaging GFP fluorescence is indicated.

4.2.3 The enhancer of the *vestigial* gene in *Apis mellifera* does not have a TAAAT motif

In the Hymenopteran *Apis mellifera*, the second thoracic segment (T2) has a pair of forewings and the third thoracic segment (T3) bears a marginally smaller pair of hindwings. A previous study has reported that the Ubx protein is expressed in both the fore and hindwing primordia of *Apis*, with the expression being slightly higher in the hindwing buds (Fig 4.3).

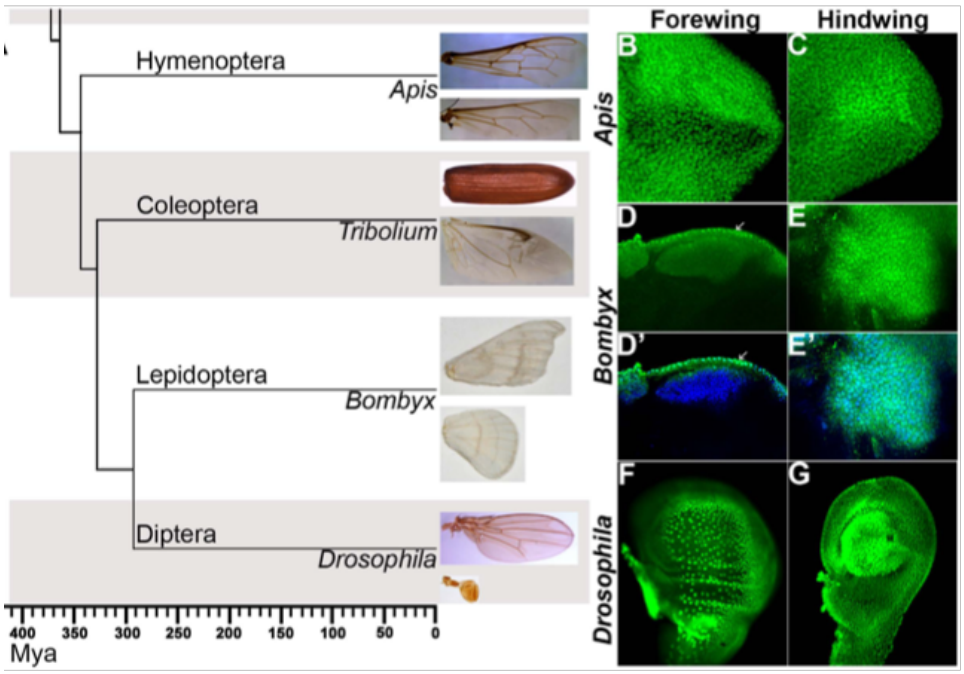


Fig 4.3: Expression of Ubx in forewing and hindwing primordia of different insect species. In *Drosophila*, Ubx is expressed only in the haltere imaginal discs. In *Apis mellifera*, Ubx is expressed in both forewing and hindwing primordia with the expression being slightly higher in the hindwings.

Genome wide comparison of Ubx targets between the *Drosophila* halteres and *Apis* hindwings suggest that a large number of genes involved in wing patterning processes are common targets of Ubx in both species. Of those common targets, a few genes are differentially regulated between *Apis* and *Drosophila*. While gene like *vestigial* (*vg*) and *wingless* (*wg*) are expressed in both the forewing and hindwing primordia in *Apis*, Ubx represses the expression of these genes in *Drosophila* halteres. Following investigation identified an orthologous

enhancer of *vg* in *Apis* (hereafter referred to as *Apis-vg*) which showed similar patterns and levels of expression between wing and haltere discs in a transgenic *Drosophila* assay (Fig 4.4). This was in contrast to the *quad-vg* enhancer which is expressed in wing imaginal discs but repressed by Ubx in the halteres (Fig 4.4).

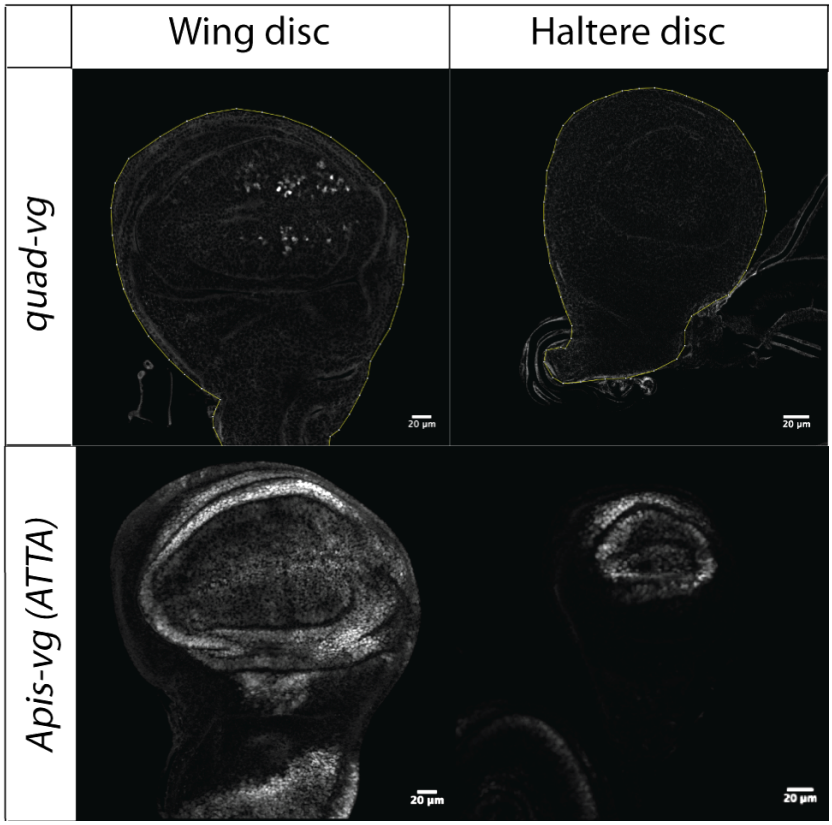


Fig 4.4: The differential regulation of enhancers of *vg* from *Drosophil*a and *Apis* by Ubx
The *quad-vg* enhancer drives reporter GFP expression in the wing imaginal discs in transgenic *Drosophila* but is repressed by Ubx in halteres. The *Apis-vg* enhancer drives a similar level of reporter GFP expression in both wing and haltere imaginal discs in transgenic *Drosophila*.

To understand this differential regulation of the *Apis-vg* and *quad-vg* enhancers by Ubx, Naveen and Tarikere et. al, 2015 focused on differences in transcription factor motif differences between the two enhancer sequences. Their findings indicate that both enhancers display a similar array of TF binding sites, except for a motif bound by the Adf1 protein which was conspicuously absent in the *Apis-vg* enhancer. However, experiments with *Drosophila*

transgenics carrying mutations in the Adf1 binding site in the *quad-vg* enhancer were inconclusive. Thus, a specific differentiating region in the two enhancers that is pertinent to understand the differential regulation of *vg* expression in *A. mellifera* and *D. melanogaster* remained to be discovered.

We revisited this problem in light of our findings about the TAAAT motif function in determining regulatory output of Ubx. Primarily we focused on differences, if any, in Ubx binding motifs between the two enhancers. We performed motif scanning analyses for the TAAAT and TAAT motifs in *Apis-vg* using the MAST software from MEME suite. This analysis revealed that within the 515bp of the *Apis-vg* enhancer, there was a single TAAT motif and no TAAAT motifs (Fig 4.5). This lack of TAAAT motif was in sharp contrast to the *quad-vg* enhancer where both TAAT and TAAAT motifs are present (Fig 4.5). Therefore it was possible that the TAAAT motif could be central to the differential regulation of the *quad-vg* and *Apis-vg* enhancers.

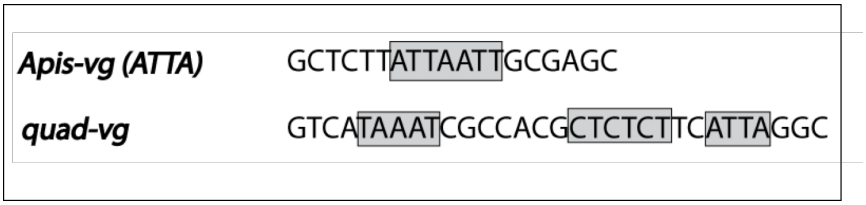


Fig 4.5: Ubx motif comparison of *Apis-vg* and *quad-vg* enhancers
Part of the *Apis-vg* and *quad-vg* enhancer containing Ubx binding motifs. While the *quad-vg* has both the TAAT and TAAAT motifs, the *Apis-vg* has only one TAAT motif and no TAAAT motifs

4.2.3 The TAAAT motif is sufficient to bring the otherwise unresponsive enhancer of *vestigial* from *Apis* under the control of Ubx in transgenic *Drosophila*

To investigate whether differences in Ubx binding motifs between the *Apis-vg* and *quad-vg* are responsible for their differential regulation by Ubx in *Drosophila*, we generated several *Drosophila* transgenics carrying the wild type *Apis-vg* enhancer and mutations in the TAAT motif of the *Apis-vg* enhancer (Fig 4.6).

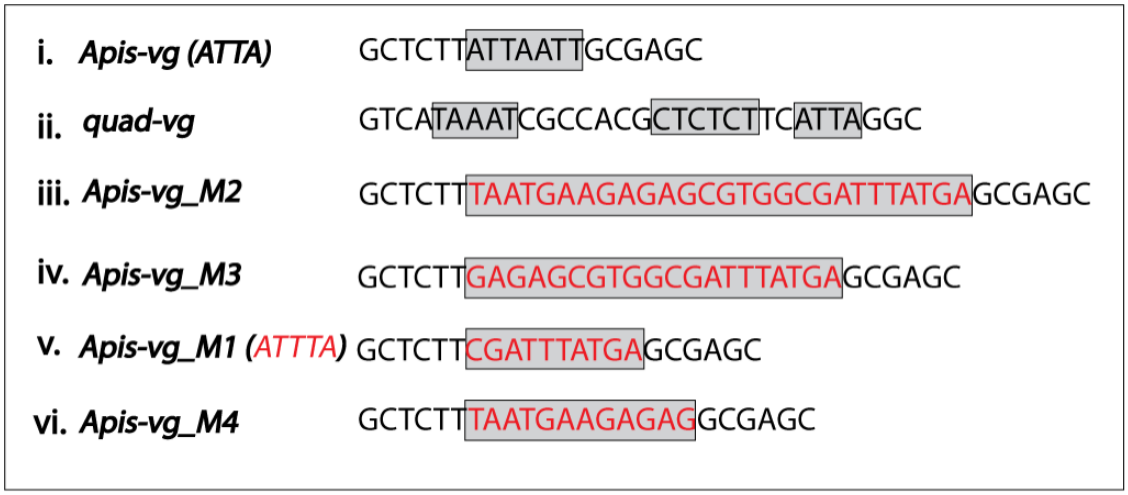


Fig 4.6: Sequence of the wild type and mutant *Apis-vg* enhancers
Part of the wild type 575bp long *Apis-vg* enhancer (i) and part of the 805bp long *quad-vg* enhancer (ii) containing the TAAT and TAAAT motifs. Mutations were generated in the *Apis-vg* enhancer by replacing the TAAT motif with various parts of the cassette in the *quad-vg* enhancer containing the TAAAT and TAAT motifs.

First, we generated a chimeric enhancer (*Apis-vg_M2*) where the TAAT motif in the *Apis-vg* enhancer was replaced by the entire 25bp cassette containing both the TAAT and TAAAT motifs from *quad-vg* enhancer. Similar to previous reports, we find that the *Apis-vg* enhancer drove strong expression of the GFP reporter in both wing and haltere imaginal discs (Fig 4.7). The *Apis-vg_M2* enhancer drove a similar level of GFP reporter expression in the wing imaginal discs. However, the GFP expression in the haltere pouch was much lower as compared to the wild type *Apis-vg* enhancer (Fig 4.7).

In another chimeric enhancer, the TAAT motif was replaced by part of the 25bp cassette containing the TAAAT motif (*Apis-vg_M3*) (Fig 4.7). This too resulted in reduced GFP expression in the haltere pouch as compared to the wild type *Apis-vg* enhancer whereas the expression in the wing imaginal discs were similar (Fig 4.7).

Interestingly, in another chimeric enhancer where the TAAT motif was replaced by part of the 25bp cassette not containing the TAAAT motif, the GFP expression in the haltere pouch was similar to that of the wild type *Apis-vg* enhancer (Fig 4.7). More importantly, changing the TAAT motif of the *Apis-vg* enhancer to the TAAAT motif (*Apis-vg-M1*) was sufficient for the significant repression of GFP reporter expression in the haltere pouch without affecting its expression in the wing pouch (Fig 4.7, Fig 4.8, Fig 4.9).

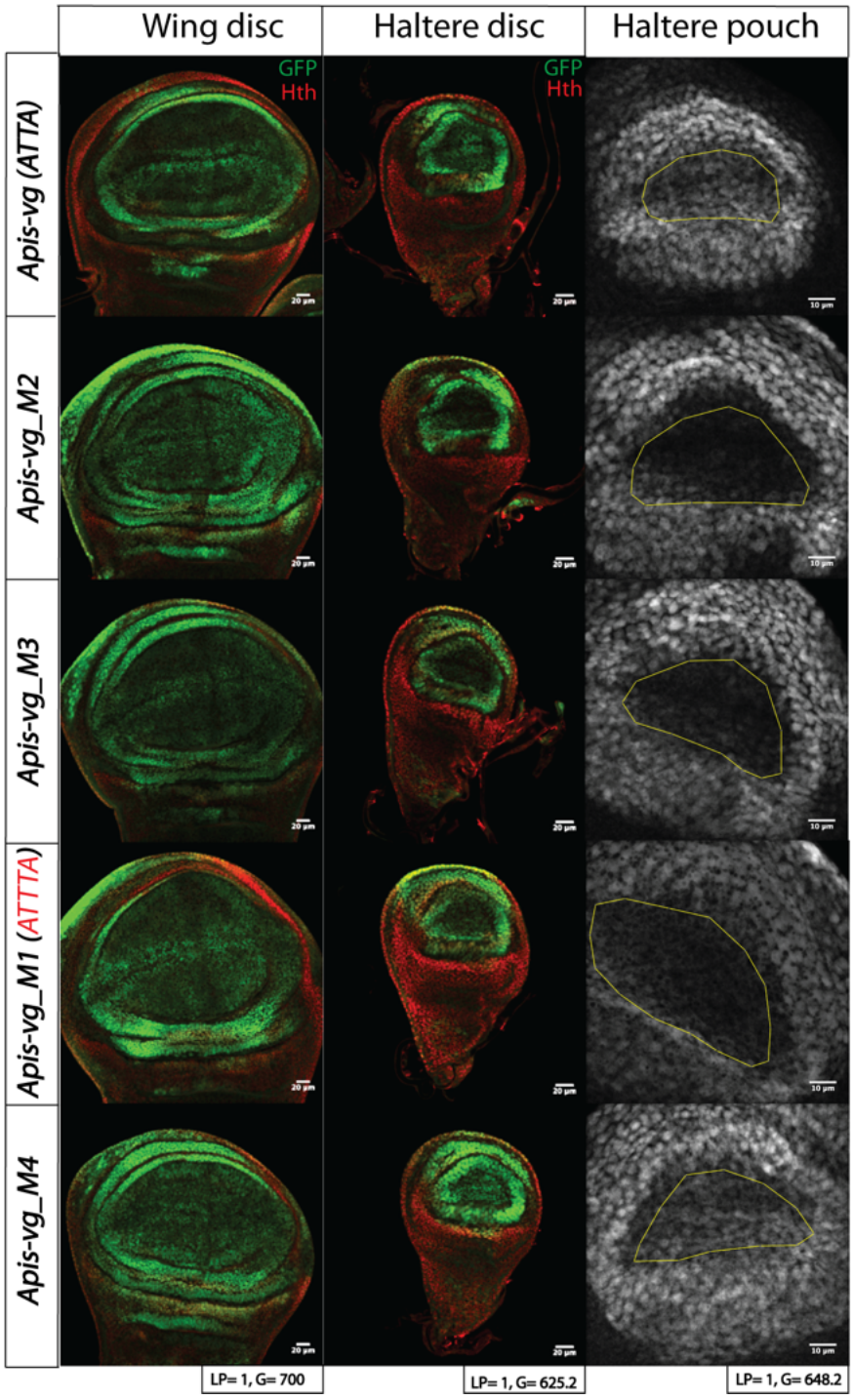


Fig 4.7: Role of TAAAT motif in Ubx mediated repression of the *Apis-vg* enhancer
The wild type *Apis-vg* enhancer drives expression of the reporter GFP in both wing and haltere imaginal discs. Mutations replacing the TAAT motif in *Apis-vg* enhancer with the parts of the 25bp cassette containing the TAAAT motif from *quad-vg* enhancer (*Apis-vg_M1*, *Apis-vg_M2*, *Apis-vg_M3*) lead to repression of reporter GFP in haltere pouch while not affecting the wing pouch. Mutations replacing the TAAT motif with part of the 25bp cassette which excludes the TAAAT motif (*Apis-vg_M4*) do not show loss of reporter expression in the haltere pouch.

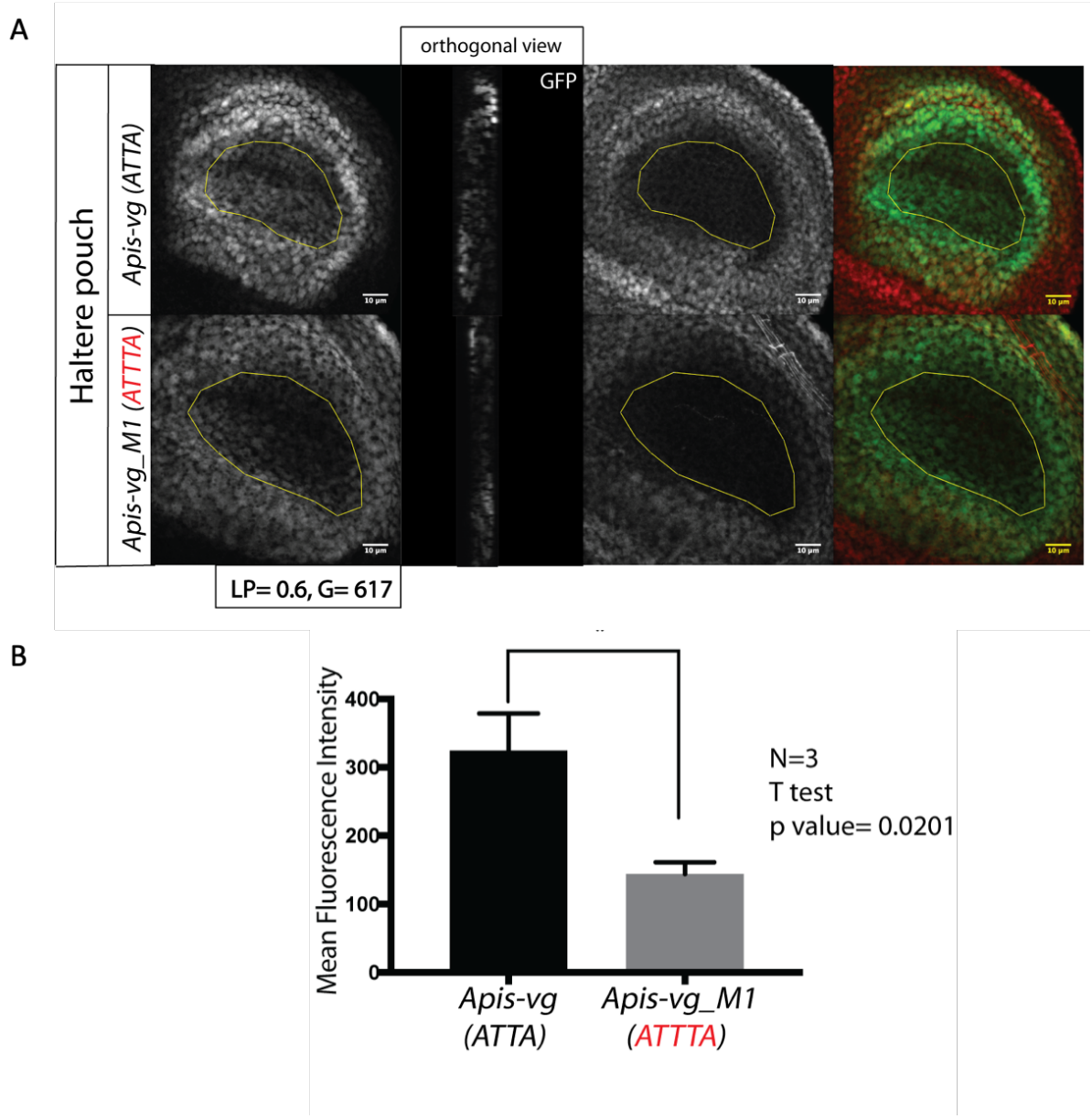


Fig 4.8: Ubx regulation of the mutant *Apis-vg* enhancer in haltere pouch

A: Magnified images of the haltere pouch of *Drosophila* transgenics expressing GFP under *Apis-vg* and *Apis-vg_M1* enhancers. GFP levels are reduced specifically in the haltere pouch of *Apis-vg_M1* transgene. Orthogonal views of the haltere imaginal pouch indicate a clear difference in GFP expression driven by the wild type and mutant enhancers. Hth staining, which is hinge-specific is used to demarcate the pouch region.

B: Quantification of average fluorescence intensity in the haltere pouch for the *Apis-vg* and *Apis-vg_M1* transgenics. Statistical test was performed using t-test (two tailed).

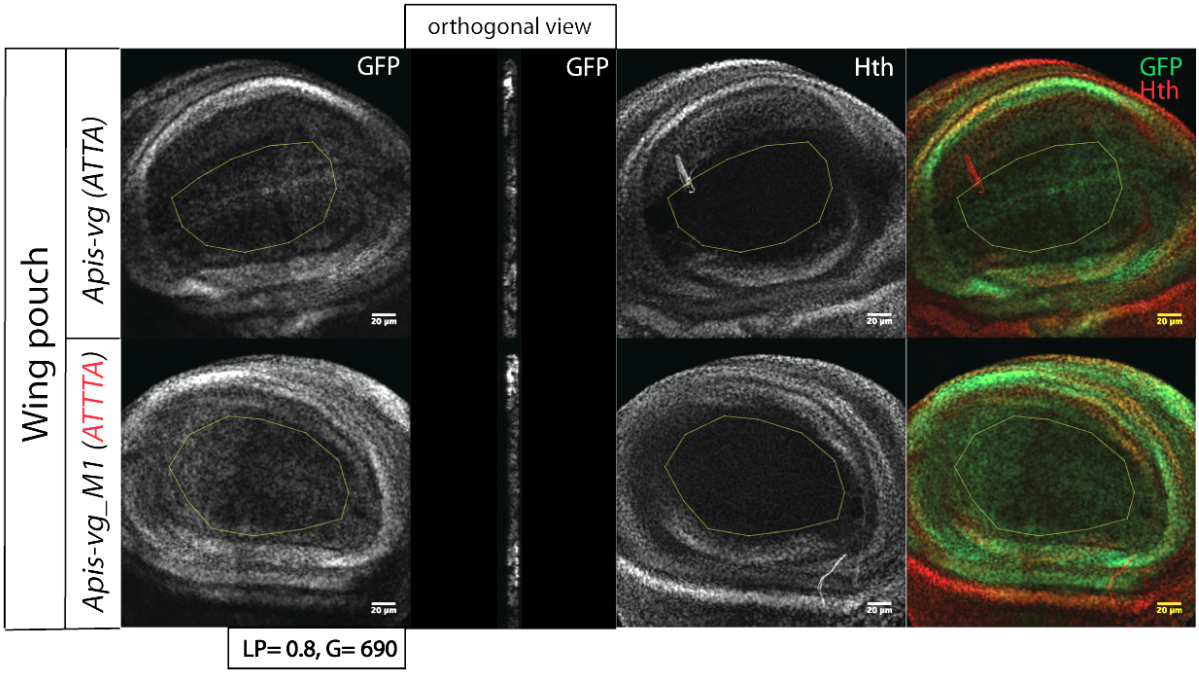


Fig 4.9: Ubx regulation of the mutant *Apis-vg* enhancer in wing pouch
Magnified images of the wing pouch of *Drosophila* transgenics expressing GFP under *Apis-vg* and *Apis-vg_M1* enhancers. Orthogonal views of the wing imaginal pouch indicate a similar level of GFP expression driven by the WT and the mutant enhancers. Hth staining, which is hinge-specific is used to demarcate the pouch region. The laser power (LP) and gain (G) used for imaging GFP fluorescence is indicated.

We also generated another chimeric enhancer where a single point mutation was introduced within the TAAT motif in the *Apis-vg* enhancer to change it to TAAAT motif. However, we did not see a significant reduction in GFP expression in the pouch region as compared to the wild type *Apis-vg* enhancer. This could possibly be attributed to chromatin architecture of the flanking nucleotides which in turn affect the efficient binding of Ubx to the TAAAT motifs. Taken together, our results suggest the role of a small change in the Ubx binding motif (TAAT to TAAAT) in the enhancer sequences of the *vestigial* gene that may have brought it under Ubx control in the dipteran lineage.

4.3 Discussion

Hox genes are known to act as master control genes for the segment-specific regulation of developmental pathways. However, owing to their highly conserved homeodomain region and their low specificity of binding in-vitro, the mechanisms by which Hox protein mediate the development of specific morphological features is largely debated. In *Drosophila melanogaster*, the Hox protein Ultrabithorax (Ubx) specifies the haltere fate by upregulating and downregulating a large number of genes involved in the wing development pathway. Identifying the mechanisms by which Ubx regulates a few key genes, thus, might help in understanding how it specifies the haltere development.

Our analysis of genome wide Ubx binding data suggest that it identifies and binds strongly to a motif containing a TAAAT core (known as the TAAAT motif). We studied the consequence of Ubx binding to the TAAAT motif in the context of regulation of two genes: *CG13222*, which is upregulated in the haltere and *vg*, which is downregulated in the haltere. Our observations suggest that the TAAAT motif is bound with higher affinity and specificity by Ubx as compared to the TAAT motif and is critical for the Ubx mediated upregulation of *CG13222* in *Drosophila* halteres. We also find that the TAAAT motif in the quadrant enhancer of *vg* (*quad-vg*) is highly conserved across most *Drosophila* species. However, we were unable to assess the role of the TAAAT motif in regulation of the *quad-vg* enhancer since most mutations in the enhancer led to loss of reporter GFP expression from both wing and haltere imaginal discs. This suggests that the region being mutated might be important for the regulation of the *quad-vg* enhancer and thus might also be indispensable for the repression of the *vg* gene by Ubx in halteres. This, however, needs to be empirically tested.

A previous report (Prasad and Tarikere et al., 2016) as well as our results suggest that an enhancer of the *vg* gene from *Apis* (*Apis-vg*) drives expression of reporter GFP in wing imaginal discs in a transgenic assay with the pattern bearing semblance to the quadrant enhancer of *Drosophila*. The *Apis-vg* enhancer drives a similar level of expression in the haltere discs and is not under the regulation of Ubx. Sequence analysis of this enhancer indicated that it has a single TAAT motif and no TAAAT motif. Transgenic assays using the *Apis-vg* enhancer allowed us to understand the role of the TAAAT motif as replacing the TAAT motif with TAAAT motif, quite appreciably, brought the enhancer under the negative regulation of

Ubx. Interestingly, the repression of reporter expression driven by the mutant enhancer was observed in the pouch region only and not in the hinge region, suggesting that manipulation of the TAAT motif to the TAAAT motif of *Apis-vg* enhancer brings its expression patterns closer to that of *quad-vg* of *Drosophila*. As *vg* is a pro-wing selector gene, its repression is critical for haltere specification by Ubx in *Drosophila*. Our studies, thus, point to a mechanism where a microevolutionary changes replacing the TAAT motif with TAAAT motif in the enhancer of *vg* might be critical for the haltere specification in the dipteran lineage.

Chapter 5

Genome wide enrichment analysis of the TAAAT motif in enhancers of Ubx targets in *Drosophila* and *Apis*

5.1 Introduction

In *Drosophila melanogaster*, the Hox protein Ubx is expressed and required for the specification of halteres. Ubx is also expressed in the hindwing primordia of other insect species (Carroll 1995; Prasad, Tarikere et al. 2016) (Fig 5.1). For example, in the Coleopteran *Tribolium castaneum*, the second thoracic segment (T2) has a pair of modified wing structures known as elytra whereas the third thoracic segment (T3) bear a pair of hindwings. Ubx is known to be expressed and required for the repression of the T2 fate and specification of hindwings in the T3 (Tomoyasu et al., 2005). In Lepidopterans like *Precis coenia*, Ubx is known to specify differences in eyespot patterns between the hindwings and forewings (Weatherbee et al., 1999). On the other hand, In Hymenopterans like *Apis mellifera* which bear a marginally smaller hindwing as compared to the forewing, Ubx is expressed in both forewing and hindwing primordia. The expression, however, is stronger in the hindwing primordia (Prasad, Tarikere et al. 2016).

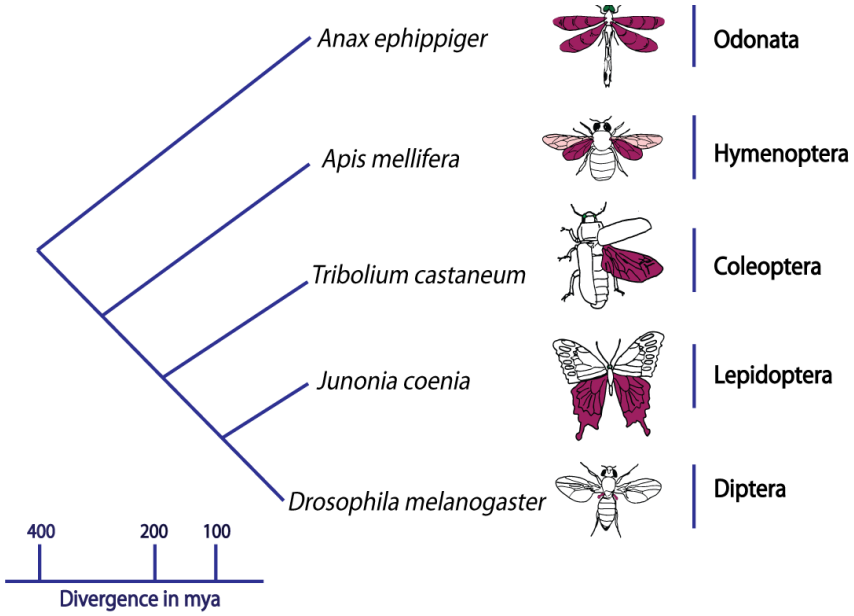


Fig 5.1: Ubx expression and hindwing morphology in insect species
The different hindwing morphologies in closely related insect species. While Ubx is required for haltere development in Dipterans such as *Drosophila melanogaster*, differences in hindwing morphology in the T3 segment of diverse insects is not correlated to the mere presence of Ubx

Overexpression of Ubx in developing wing imaginal discs leads to suppression of the wing fate and specification of the haltere fate in T2. Interestingly, overexpression of Ubx derived from *A. mellifera or B. mori* or *T. castaneum* in developing wing imaginal discs also causes wing to haltere transformation, suggesting that changes at the protein level in Ubx may not have contributed significantly to the diversification of structures between different species. Additionally, comparison of genome-wide targets of Ubx between developing halteres in *Drosophila* and developing hindwings of *Apis mellifera* and *Bombyx mori* species (Prasad, Tarikere et al. 2016) suggest that a large number of wing patterning genes constitute the common targets of Ubx between the three species. Only a few of these targets are differentially expressed between the developing wing and haltere in *Drosophila* but not between the developing forewing and hindwing in *A. mellifera* or *B. mori*. This suggests that evolution of Ubx function might be a consequence of changes in the cis-regulatory sequences of common targets.

Supporting this hypothesis, our observations indicated that a motif with a TAAAT core (known as the TAAAT motif) is enriched at genome wide Ubx binding sites in the developing halteres in *Drosophila*. Furthermore, the TAAAT motif was bound with higher affinity and specificity, in-vitro. And a combination of *in vitro* and *in* vivo assays revealed necessity of the TAAAT motif for the Ubx mediated activation of the CG13222 gene as well as its sufficiency to bring an otherwise unresponsive enhancer of the vg gene from *Apis* under Ubx control. Thus, the differences in Ubx binding motif (TAAAT vs TAAT) in enhancers of orthologous targets may have brought key wing patterning genes like *vg* under the differential regulation of Ubx in the dipteran lineage. We wanted to understand whether such differences in Ubx binding motif are a recurrent theme which contributes to the differential hindwing morphology. To test this, we performed extensive analyses comparing the Ubx protein between the *Apis* and *Drosophila* as well as its genome wide binding studies for Ubx in *Drosophila* and *Apis*. Interestingly, the TAAAT motif was enriched by Ubx specifically in *Drosophila* and not in *Apis mellifera*, even though we did not find any differences between the DNA binding domain of the two Ubx proteins. On the other hand, binding motifs for transcription factors like GAGA, brk and CTCF were enriched in both datasets. Additionally, we observed a difference in the clustering of Ubx binding sites between the two species with Ubx binding closer to the TSS in *Apis* but not in *Drosophila*. Taken together, our studies point to a role of microevolutionary changes in the binding motif of Ubx (TAAT to TAAAT) as a plausible mechanism for morphological evolution of hindwings in dipterans and hymenopterans.

5.2 Results

5.2.1 The DNA binding domain of Ubx is highly conserved between *Drosophila* and *Apis*

Differences in hindwing morphology between *Apis* and *Drosophila* can be a result of various factors like a) differences in Ubx expression, b) differences in protein sequence, c) differences in Ubx targets or d) differences in cis-regulatory sequences of common target genes. Previous studies have indicated that Ubx is indeed expressed in both developing halteres in *Drosophila* as well as developing hindwing in *Apis mellifera* (Fig 5a), suggesting that Ubx expression levels contribute marginally to the morphological divergence.

We next carried out sequence comparison of the Ubx protein derived from various insect species like *Drosophila melanogaster*, *Apis mellifera*, *Bombyx mori*, *Junonia coenia and Tribolium castaneum*. Similar to previous reports, our analysis indicated that the DNA binding homeodomain region is highly conserved across all species (Fig 5b). We also observe high level of conservation for the Hexapeptide (HX) and Ubda motifs which are used by Ubx for interaction with TALE (Three amino acid loop extension) group of proteins like Extradenticle (Exd).

In contrast, the linker region in *Drosophila* Ubx is prominent compared to other species (Fig 5.2). While the linker region spans a length of approximately 38 amino acids in *Drosophila* Ubx, it is almost inconspicuous in other species. Interestingly, a variant of Ubx which does not contain the linker region, Ubx IVA, is unable to specify the haltere fate in T3 in transgenic *Drosophila* (De Navas et al., 2011). Additionally, the linker region is also important for interaction of Ubx with co-transcription factors like Exd via the HX motif (Passner et al., 1999). This suggests that a shorter linker region in other insects like *Apis* can affect the Ubx-cofactor mediated regulation of target genes in *Apis*. To this end, we tried to model the interaction of known cofactors of Ubx like MAD and Exd with Ubx from different species. While our analyses predict that length of the linker region can play an important role in HX motif mediated Ubx-cofactor interactions, the contribution of the linker region in causing Ubx mediated morphological differences between *Drosophila* and *Apis* still needs to be empirically determined.

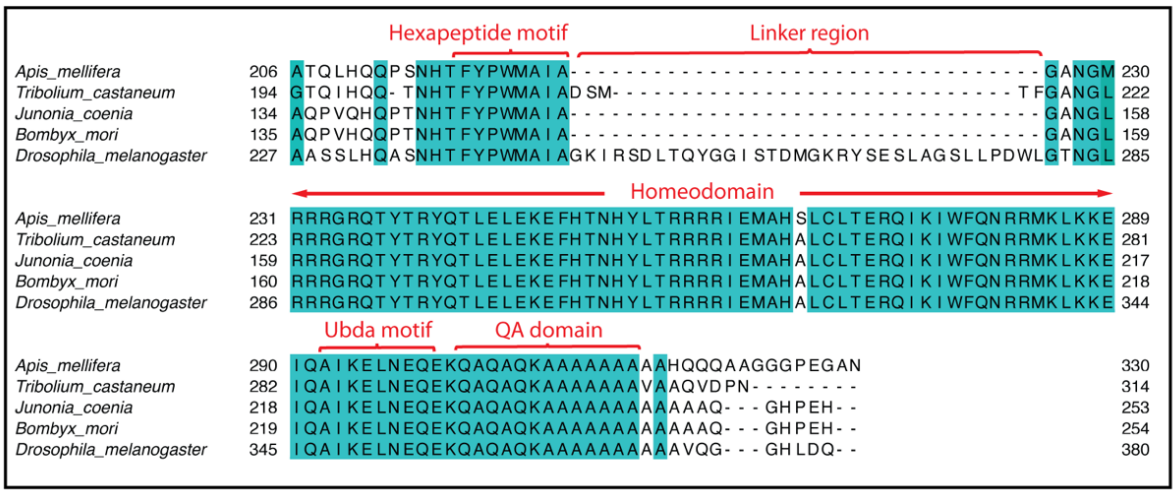


Fig 5.2: Ubx protein comparison across different insect species
Comparison of Ubx protein sequences between the Hymenopteran (*Apis mellifera*), Coleopteran (*Tribolium castaneum*), Lepidoptera (*Junonia coenia* and *Bombyx mori*) and Dipteran (*Drosophila melanogaster*). The DNA binding homeodomain region and protein interaction motifs, hexapeptide and Ubda, are highly conserved across all species. The linker region of Ubx is inconspicuous in all species other than *Drosophila*.

5.2.2 Wing patterning genes have remained common targets of Ubx in *Drosophila* and *Apis*

We next wanted to investigate whether differences in Ubx target genes between *Drosophila* and *Apis* can throw light into differences in their hindwing morphology. Using the Ubx ChIP-seq data for *Apis* hindwings (generated by Prasad, Tarikere et al. 2016) and for *Drosophila* halteres (generated in this study), we analyzed Ubx bound regions in *Drosophila* and *Apis* genome and identified direct targets by annotating Ubx bound peaks to the nearest TSS in each species. Next, we performed gene ontology (GO) analyses of Ubx targets. Consistent with previous study, our analysis suggested that wing patterning genes have remained common targets of Ubx in both insect species (Fig 5.3). A few of these genes, notably *vestigial, spalt* and *wingless*, are known to be differentially expressed between the wing and haltere in *Drosophila* but not in the forewing and hindwing primordia in *Apis*. This indicated differences in regulation of target genes between *Drosophila* and *Apis* may be attributed to differences in cis-regulatory regions of common targets, which further impinge on differences in hindwing morphology.

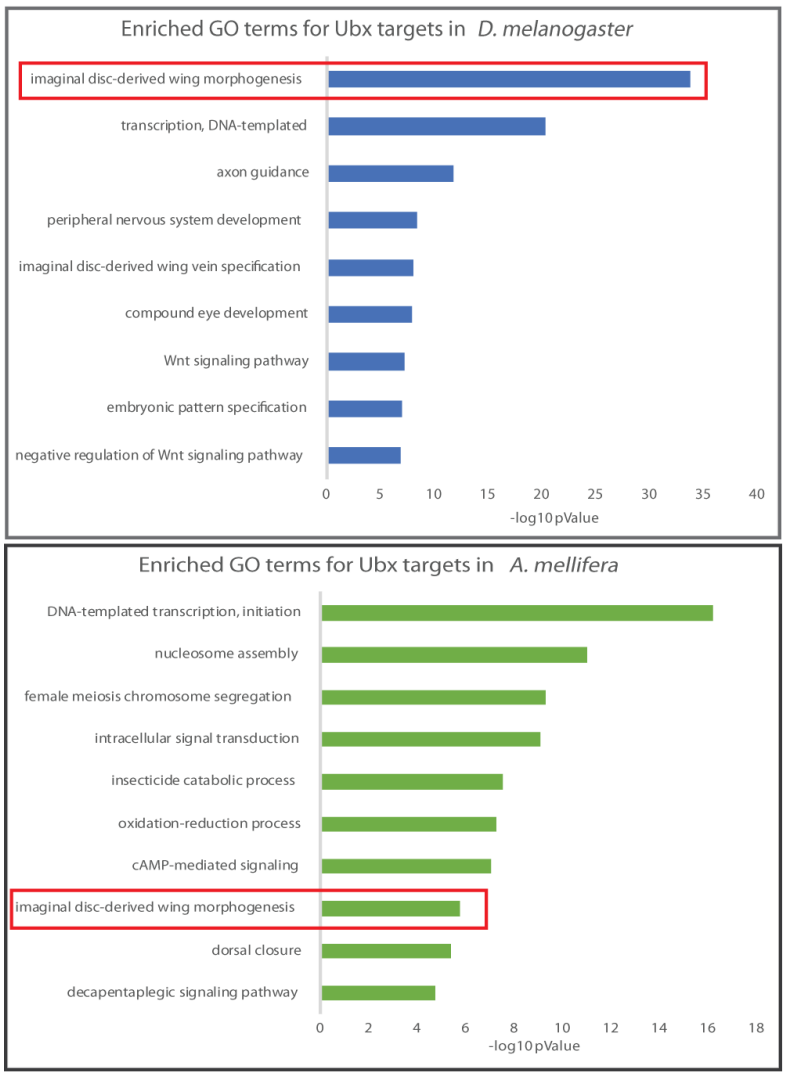


Fig 5.3: Genome wide comparison of Ubx targets in *Apis* and *Drosophila*
Gene ontology analyses of direct targets of Ubx indicate that wing patterning genes are targeted by Ubx in both *Apis* hindwings and *Drosophila* halteres. The proportion of wing patterning genes, however, is larger in *Drosophila* as compared to *Apis*.

5.2.3 The high affinity TAAAT motif in enhancers of Ubx targets is specific to *Drosophila melanogaster*

We wanted to understand the differences in cis-regulatory sequences of Ubx targets which lead to their differential regulation in *Drosophila* and *Apis*. To achieve that, we first sought to identify differences, if any, in transcription factor binding profiles around Ubx bound regions in *Drosophila* and *Apis*. We performed de-novo motif analyses of Ubx pulled down regions from both species on a 200bp region centered around the peak summit. In contrast toto

enrichment of TAAAT core (known as the TAAAT motif) in *Drosophila* halteres, we did not observe enrichment of the TAAAT motif in *Apis* hindwings (Fig 5.4). Additionally, our results reveal that the TAAT motif is not enriched by Ubx in either *Drosophila* or *Apis* dataset, however, binding sites for other transcription factors like CTCF, GAGA factor and Brk are enriched in both datasets (Fig 5.4).

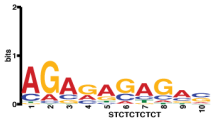
Enriched motif from de-novo analysis		p Value	
		D. melanogaster	*A. mellifera*
	TAAAT motif	1e-113	Not enriched
	CTCF	1e-39	1e-431
	Trl/GAGA	1e-60	1e-250
	Brk	1e-17	1e-17

Fig 5.4: De-novo motif analysis comparison of Ubx pulled down sequences
The TAAAT motif is enriched by in Ubx pulled down sequences in *Drosophila* halteres but not in *Apis mellifer*a hindwings. The TAAT motif is not enriched either dataset. Binding motifs for transcription factors like CTCF, GAGA and Brk are enriched in both datasets.

Our analyses revealed that the TAAAT motif is enriched by 1.7folds in Ubx pulled down sequences in *Drosophila* as compared to the whole genome (Fig 5.5). In contrast, the frequency of the TAAAT motif was 1.8folds less as compared to the whole *Apis* genome. Additionally, the frequency of the TAAAT motif in *Apis* genome was 1.8folds higher than its frequency in the *Drosophila* genome (Fig 5.5).

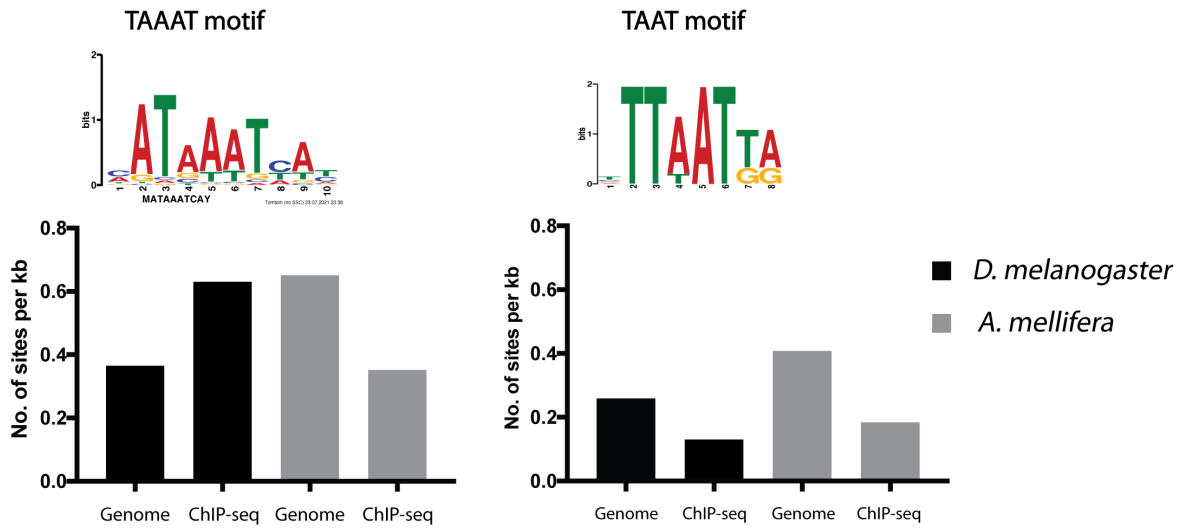


Fig 5.5: Enrichment of the TAAAT and TAAT motif in enhancers of Ubx targets in *Drosophila* and *Apis*

The frequency of the TAAAT motif is 1.7folds more in Ubx pulled down sequences in *Drosophila* as compared to the whole genome. In *Apis*, the frequency of the TAAAT motif is 1.8folds less as compared to the frequency in the whole genome. The TAAT motif is not enriched in either of the datasets

Furthermore, we also observed differences in clustering of Ubx binding sites in *Drosophila* and *Apis*. While Ubx binding sites were clustered away from the TSS in *Drosophila*, the reverse is observed for *Apis mellifera*. However, whether the difference in topology of binding sites holds a functional significance in the context of morphological diversification by Ubx needs to be tested empirically. Taken together, in the context of the differential specification of the T2 and T3 segment in *Apis* and *Drosophila* by Ubx, our results help uncover a mechanism in which micro-evolutionary changes in the Ubx binding sites might have brought key wing patterning genes under the control of Ubx in the dipteran lineage.

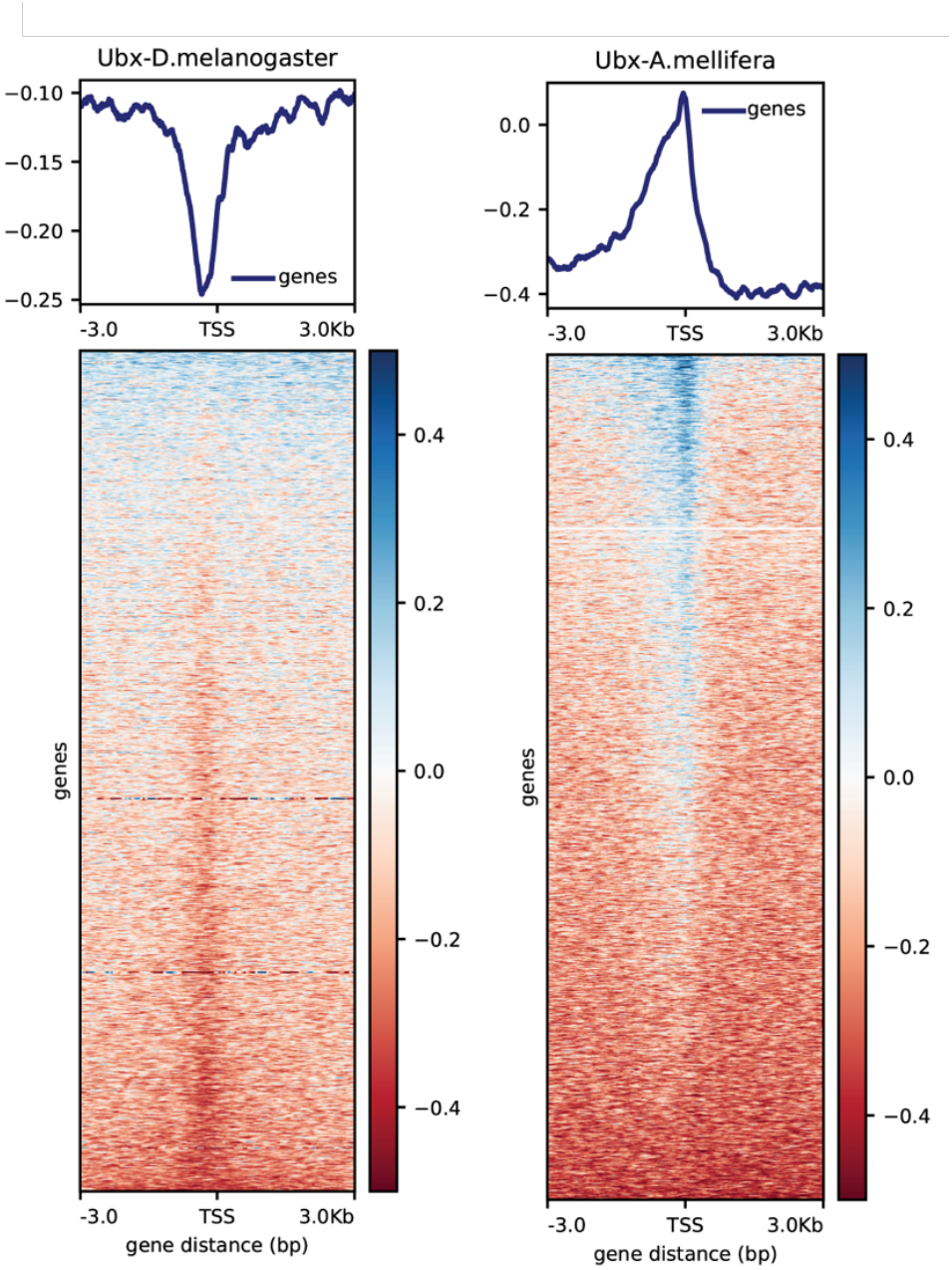


Fig 5.6: Topology of Ubx binding in *Drosophila* and *Apis*
Ubx binding sites are clustered away from the TSS in *Drosophila melanogaster* halteres. The reverse is observed for *Apis mellifera* hindwings.

5.3 Discussion

The role of Hox genes in the segment-specific development of an organism along the anterior-posterior axis, has been established quite clearly by studies carried over the past few decades. In addition, divergence in Hox gene number, expression and function is directly correlated to diversification of body plan in the metazoan lineage. However, the fact that homologous Hox genes specify fundamentally different body plans in different species has baffled biologists for a long time. While it is now widely accepted that such differences in morphology are to some extent, but not exclusively, a consequence of differential regulation of a set of common targets, the exact mechanisms behind such differential regulation are not well understood.

Our studies comparing the genome wide targets of Ubx between *Drosophila* and *Apis* suggest that a large number of wing patterning genes are common targets of Ubx in both species. Previous studies have indicated that some of these targets like *spalt, wingless* and *vestigial (vg)*, are differentially expressed between the wing and halteres in *Drosophila* but not between fore and hindwing in *Apis*. Results using transgenic *Drosophila* reveal that a microevolutionary change in the Ubx binding motif (TAAT to TAAAT), was sufficient to repress the otherwise unresponsive enhancer of the *vg* gene from *Apis*. We also observed, that in general, the TAAAT motif is more prevalent in Ubx pulled down sequences in *Drosophila* halteres but not in the *Apis* hindwings. This suggests that changes in the cis-regulatory regions, specifically in Ubx binding motifs which confer stronger binding, might have brought certain important wing patterning genes like *vg* under Ubx regulation in the dipteran lineage. Since genes like *vg* control important regulatory networks required for the wing specification process, the modulation of such genes by Ubx would in turn lead to changes in the chromatin landscape of many other targets. This would in turn make them sensitive to Ubx mediated regulation, thereby facilitating the haltere specification process.

In addition, the Ubx protein in *Drosophila* has a larger linker region between the protein interaction motif (HX) and the DNA binding domain, than in *Apis*. We propose that a longer linker region might facilitate stronger binding of Ubx to the chromatin by interaction with other transcription factors present in the vicinity. Once bound strongly, the Ubx protein might influence other proteins in the neighborhood, thereby regulating the gene expression in halteres. Thus, evolutionary changes in the protein structure as well as Ubx binding motif (TAAT to TAAAT) in enhancers of common Ubx targets might both be factors involved in the evolution of Ubx function.

Chapter 6

Mechanisms governing recognition of activated vs repressed genes by Ubx

6.1 Introduction

Our efforts to identify direct targets of Ubx and the sequences which it binds to, in-vivo, in *Drosophila* halteres were aimed to address two broad questions; 1) the mechanisms by which it selects its targets 2) the mechanisms by which it recognizes which targets to upregulate and which to downregulate. Bioinformatics analyses combined with molecular biology and genetic approaches, revealed that a motif with a TAAAT core site (known as the TAAAT motif) is functionally relevant for Ubx mediated selection of targets as well as their regulation in *Drosophila* halteres. The TAAAT motif was not only critical for the activation of the edge enhancer of the *CG13222* gene, but also instrumental in repression of an orthologous enhancer of *vg* form *Apis mellifera* in transgenic *Drosophila*. This confirmed that the TAAAT motif is used for both activation and repression of targets by Ubx in *Drosophila* halteres, thereby suggesting that the TAAAT motif might be used for stronger binding of Ubx to the chromatin and the upregulation or downregulation of target genes might be governed by other transcription factors in the vicinity of Ubx binding.

In this context, we wanted to identify and/or differentiate between molecular topography around Ubx binding sites in the haltere which may be important for the mechanism of identifying activated vs repressed targets (henceforth termed as target recognition) by Ubx in *Drosophila* halteres. To do this, we first identified enhancers of target genes that are upregulated and downregulated by Ubx. Next, we developed a GUI based analysis pipeline to calculate the frequency of each of the *Drosophila* transcription factors in upregulated and downregulated enhancers. Using this analysis pipeline, we identified a motif, predicted to be bound by the Mes2 (mesodermally expressed 2) transcription factor, which was specifically enriched in upregulated enhancers. In-vitro assays using S2 cells suggested that the Mes2 binding motif was present in close vicinity to the Ubx binding motif and was important for the activation of the *edge* enhancer of the *CG13222* gene. Additionally, using our analysis pipeline we were able to confirm the importance of the grain transcription factor, a candidate cofactor of Ubx in *Drosophila* halteres, in facilitating activation of important target genes by Ubx. Taken together, our studies not only provide a user interactive pipeline for identifying and comparing transcription factor binding occupancy between groups of enhancers, but also help throw some light onto the mechanisms of target recognition by Ubx in *Drosophila* halteres.

6.2 Results

6.2.1 Identification of differentially expressed genes between wing and haltere imaginal discs

To identify enhancer regions of genes that are upregulated and downregulated by Ubx in the *Drosophila* halteres, we first sought to identify genes that are differentially expressed between the third instar wing and haltere imaginal discs using RNA sequencing. Wing and haltere imaginal discs were isolated at 4° centigrade, transferred to Trizol solution, snap frozen and shipped to Genotypic Solutions, Bangalore for further processing. Raw fastq reads were aligned to the *Drosophila* genome (dm6) and the edgeR software used to identify differentially expressed genes between the two tissues. Three independent biological replicates were used for analysis and only those genes were considered to be expressed which had a count per million (CPM) value greater than 1 in all three replicates. Using a 1.5-fold difference as the cutoff, we identified 633 genes that are differentially expressed between the wing and halteres. Of these, 385 genes are upregulated in the haltere, whereas 248 genes are downregulated in the haltere. We compared the list of differentially expressed genes to Ubx direct targets in the haltere and identified 64 Ubx targets that are upregulated in the haltere and 33 targets which are downregulated in the haltere (Fig 6.2).

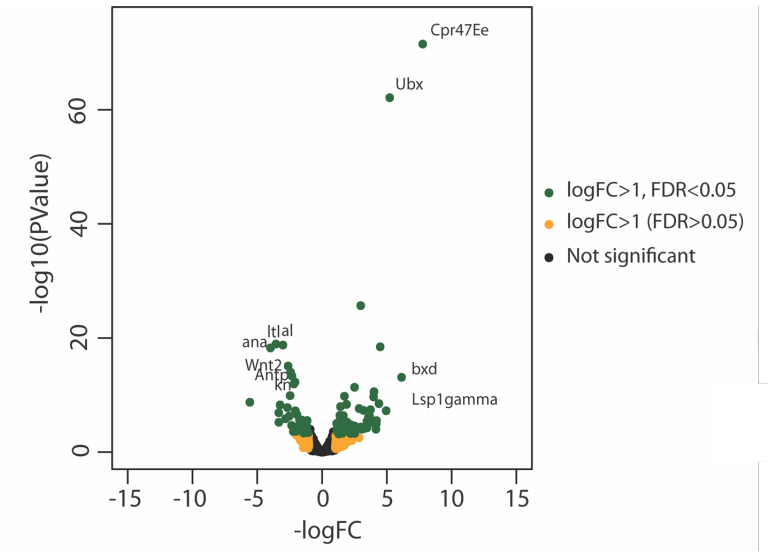


Fig 6.1: Differentially expressed genes between wing and haltere imaginal discs

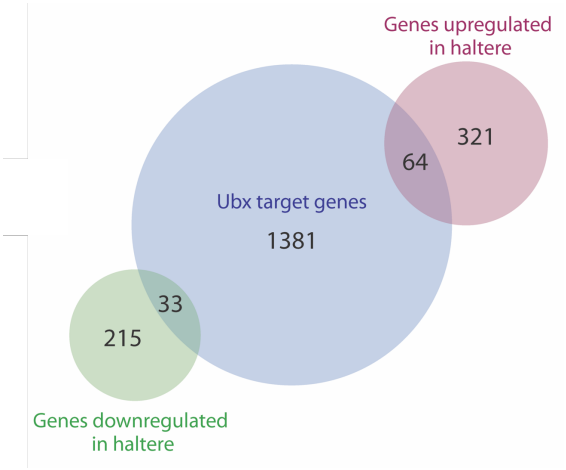


Fig 6.2: Comparison of differentially expressed genes between the wing and haltere imaginal discs to Ubx direct targets identified using ChIP-seq

6.2.2 An analysis pipeline to identify cofactor binding differences in putative enhancers of Ubx

We hypothesized that Ubx might be using different set of transcription factors for targets that are upregulated or downregulated in the haltere. Thus, understanding differences in transcription factor profiles of enhancer sequences of direct targets of Ubx is key to understand the mechanism of target selection. In this context, we developed an analysis pipeline to compute transcription factor occupancy differences between groups of enhancer sequences (Fig 6.3). We compared direct targets of Ubx from ChIP-seq to differentially expressed genes between the wing and haltere and assigned them to one of three categories; genes which are upregulated by Ubx, genes which are downregulated and genes which are not-differentially expressed between the wing and haltere. We then identified Ubx bound regions which were annotated to these genes and categorized them as Ubx response elements that are upregulated, downregulated or not-differentially expressed and scanned for transcription factor motifs in each of the categories (Table 6.1).

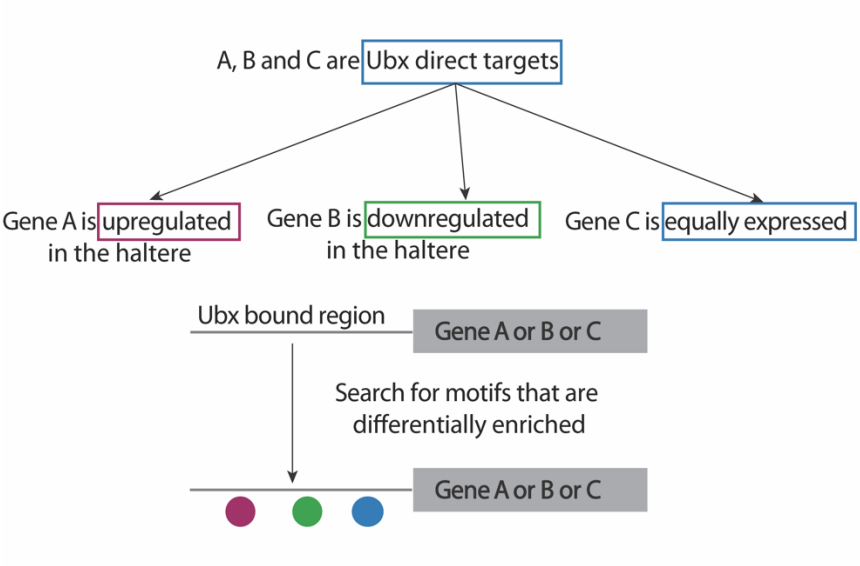


Fig 6.3: Analysis pipeline to identify TFs involved in target recognition by Ubx
An in-silico analysis pipeline to identify differences in transcription factor occupancy profiles in enhancers of genes that are upregulated or downregulated by Ubx. Comparing direct targets of Ubx with the list of differentially expressed genes between the wing and haltere help identify genes which are upregulated or downregulated by Ubx in the halteres. The Ubx bound regions of such genes are further identified from ChIP-seq dataset and categorized as upregulated, downregulated or not-differentially expressed enhancers. Motif analysis of each category of enhancer can provide insights into putative transcription factors which can assist in target recognition by Ubx in *Drosophila* halteres

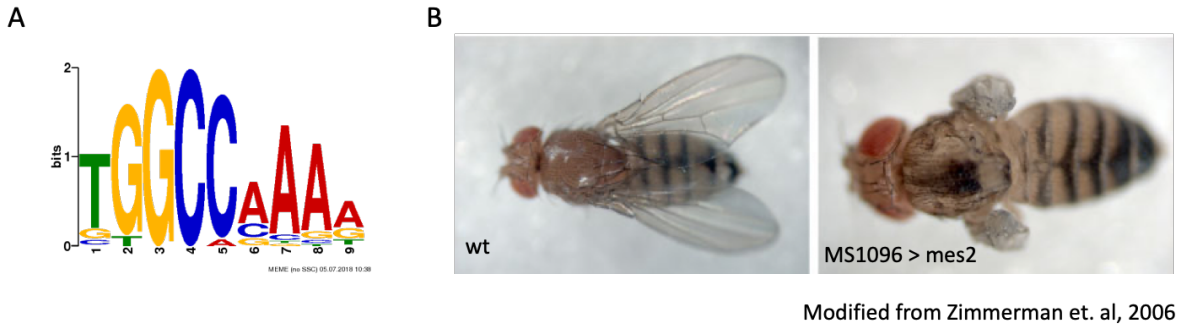
Peak numbers for various comparison groups	
Total peaks from Ubx ChIP-seq	2142
Peaks mapping to upregulated genes in haltere	97
Peaks mapping to downregulated genes in haltere	62
Peaks mapping to not-differential genes in haltere	198

Table 6.1: Categorization of Ubx bound enhancers
Table showing the number of peaks which are characterized as upregulated, downregulated or not-differentially expressed between the wing and haltere

2.2.4 A motif predicted to be bound by the Mes2 protein is enriched in putative enhancers of targets upregulated by Ubx

Next, using our automated pipeline, we scanned for transcription factor motifs in each of the enhancer categorized; upregulated, downregulated and not-differentially expressed. Our results reveal that binding sites for TFs like Tramtrack (ttk) and Mesodermally expressed 2 (Mes2) were found to be enriched in upregulated enhancers. For enhancers that are downregulated by Ubx, our studies revealed enrichment of binding site for TFs like Nubbin (nub), CTCF and Brinker (brk) (Table 6.2).

Of particular interest was the overrepresentation of Mes2 binding motif (Fig 6.4 A) in upregulated enhancers. Previously, comparative MEME analyses had also indicated a significant enrichment of the Mes2 binding motif in enhancers of upregulated Ubx targets. Additionally, previous studies have reported that overexpression of Mes2 in developing wing imaginal discs leads to disruption of wing structures and its partial transformation towards haltere fate. We thus wanted to understand the possible role of this motif in Ubx mediated regulation and recognition of targets in the haltere.



Modified from Zimmerman et. al, 2006

Fig 6.4: Putative role of Mes2 in aiding target recognition by Ubx

A: The Mes2 binding motif

B: Overexpression of Mes2 in developing wing discs lead to disruption of proper wing development

Next, we sought to evaluate the importance of Mes2 in Ubx mediated regulation of target genes. The *CG13222* gene is activated by Ubx in the haltere imaginal discs. It is known to bind to a proximal enhancer (*edge*) of the *CG13222* gene located 350bp upstream of the TSS. This enhancer is also known to be activated in S2 cells on Ubx induction as suggested from Luciferase assays (explained in detail in Materials and Methods). Motif scanning analysis using the MAST software indicate the presence of a Mes2 binding motif 50bp upstream of the Ubx binding site. We generated Luciferase assay constructs of the *CG13222* enhancer with mutations in the Mes2 binding site. While the wild type enhancer of *CG13222* is activated in S2 cells by a factor of 3 folds on Ubx induction, we observe a significant reduction in activation of the mutant enhancer (Fig 6.5), thus, suggests that the Mes2 binding motif is important for the Ubx mediated activation of the *CG13222* enhancer in S2 cells.

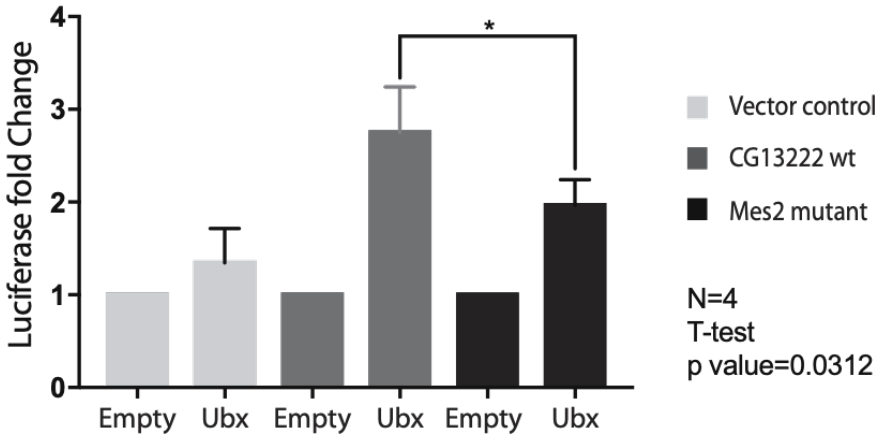


Fig 6.5: Role of Mes2 in Ubx mediated upregulation of *CG13222* in S2 cells
Luciferase assays reveal that the *CG13222* enhancer is activated by Ubx in S2 cells. Mutation of Mes2 binding motif lead to significant reduction of activation of the enhancer.

Next, to understand the role of the Mes2 binding motif in target selection by Ubx in haltere imaginal discs, we generated *Drosophila* transgenics carrying mutations in the Mes2 binding motif of the *CG13222* enhancer cloned upstream of a GFP reporter (Mes2_M). The wild type *CG13222* enhancer drives expression of the GFP reporter in posterior edge of the

haltere discs (Fig 6.6 A). However, we observed an identical phenotype in halteres of *Drosophila* transgenics carrying the mutant construct (Fig 6.6 B). We further checked the effect of Mes2 downregulation on the haltere phenotype in a *Drosophila* genetic background expressing a single allele of the Ubx gene ((abxpbxbx/+). Since the genetic background used in the study expresses the Ubx protein at a lower level compared to the wildtype, we hypothesized that knocking down any critical co-transcription factor of Ubx will drive a haltere to wing phenotype in the mutant *Drosophila*. While we observed a mild wing to haltere transformations phenotype on downregulating Mes2 in the Ubx sensitized background (Fig 6.7 B), the role of Mes2 in Ubx mediated regulation and recognition of target gene still remains to be empirically validated using extensive genetic and molecular biology methods.

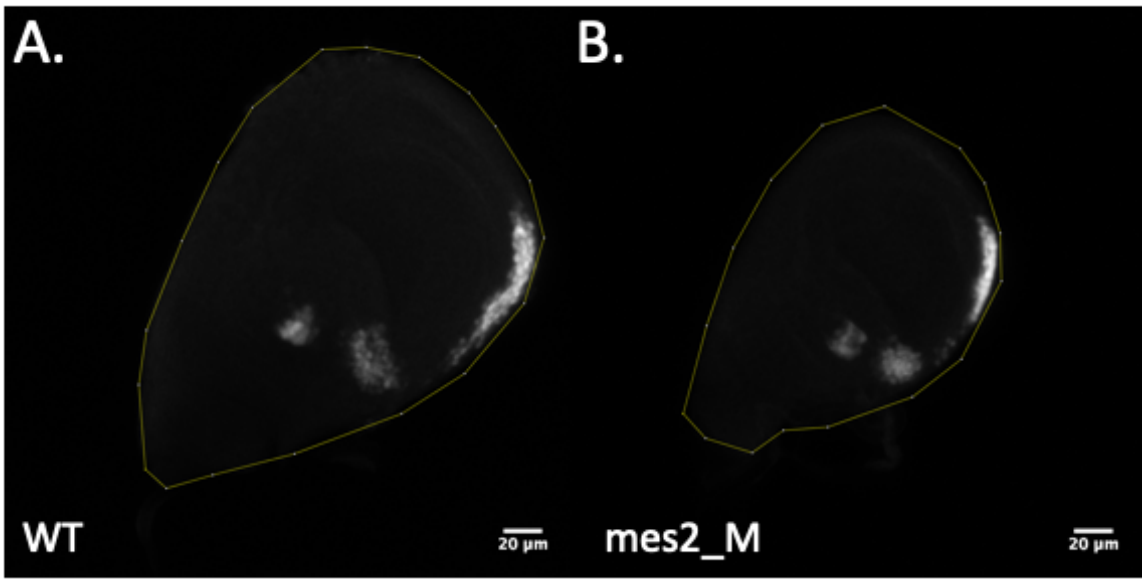


Fig 6.6: Role of Mes2 in Ubx mediated upregulation of *CG13222* in halteres
The wild type *CG13222* enhancer drives reporter GFP expression in the posterior edge of the haltere imaginal discs. The Mes2_M mutant where the Mes2 binding motif is mutated does not show any difference in reporter GFP expression.

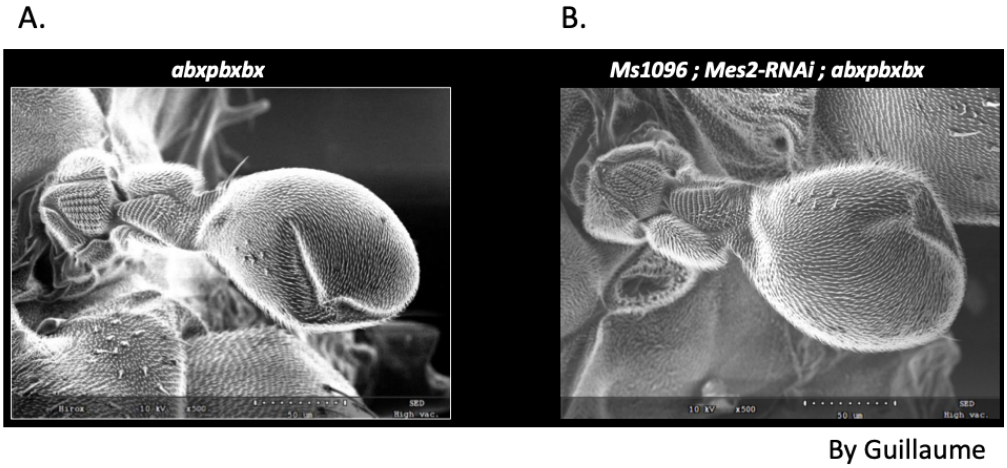


Fig 6.7: Effect of Mes2 downregulation on haltere morphology in Drosophila
The adult haltere in a Ubx sensitized (abxpbxbx/+) genotypic background (A). The haltere size is seen to increase on loss of Mes2 in developing halteres in Ubx sensitized background (B)

2.2.5 Grain as a putative transcription factor involved in the Ubx mediated downregulation of target genes

A recent screen performed by Giraud and Paul et. al, 2021, identified transcription factors which are expressed in the wing and haltere discs of *Drosophila* and are critical for the development of the halteres in the third thoracic segment. Giraud and colleagues used two genetic backgrounds, wild type and abxpbxbx/+ (Ubx sensitized background), to screen the the homeotic transformation phenotype (haltere to wing transformation) upon RNAi mediated depletion of a TF. The rationale behind this screen was that any transcription factors whose downregulation shows a homeotic transformation in the wild type or abxpbxbx/+ background might be an important candidate for Ubx mediated haltere specification process. Of the 117 TFs that were tested, their data suggests that only 7 showed phenotype in the wild type background whereas 14 showed phenotypes in the abxpbxbx/+ background. One of the transcription factors which showed haltere to wing transformation in the abxpbxbx/+ background was the Grain transcription factor. Loss of Grain driven by the MS1096 gal4 driver in a Ubx reduced background led to increase in size of the haltere, as well as increase in number of bristles in the haltere capitellum (Fig 6.8).

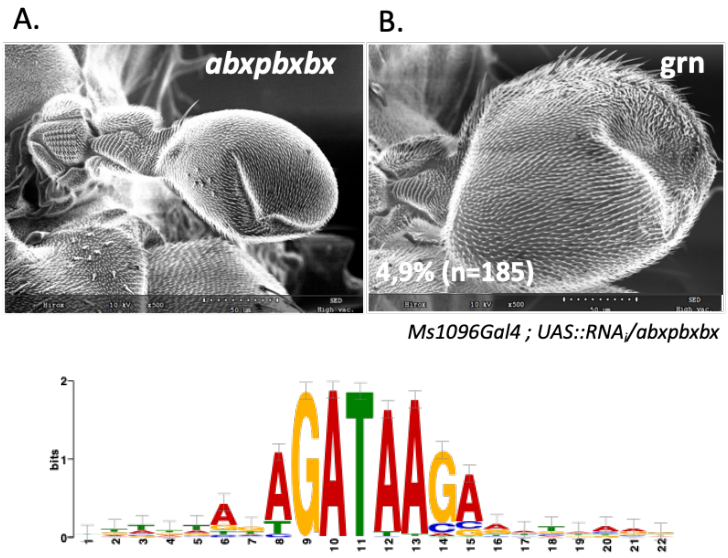


Fig 6.8: The putative role of Grain in aiding Ubx mediated downregulation of targets

A. The adult haltere in a Ubx sensitized (abxpbxbx/+) genotypic background (A). Loss of grain in developing halteres lead to severe haltere to wing transformations including increase in size and bristles of the haltere capitellum.

B. DNA binding motif of grain

We used our analysis pipeline to identify enhancer regions which have binding sites for the Grain transcription factor. Interestingly, we found that frequency of grain binding sites in downregulated enhancers was 1.5 times more than frequency in upregulated enhancers (Table 6.1). Further analysis revealed that the 15 enhancers that have grain binding motifs correspond to 12 target genes (Table 6.1). Of these 12 genes, there were critical regulators of wing development like *vg*, *wnt2*, *Dll*, *al* and *Antp*, which are repressed by Ubx in the haltere imaginal discs (Table 6.2). The functional role of the Grain motif, however, in the context of Ubx mediated downregulation of target genes still needs to be empirically tested. Taken together, our analysis pipeline provides a robust framework to verify, in-silico, the role of putative transcription in a given enhancer sequence. Additionally, this pipeline can be used to identify critical leads in solving biological questions using a whole genome sequencing approach.

Category	Frequency of Grain binding sites	Number of enhancers having Grain binding sites	Number of enhancers having both Grain and Ubx binding sites	Number of direct target genes
Upregulated in the haltere	0.193154601	13	8	13
Downregulated in the haltere	0.301146127	15	8	12
Not-differentially expressed	0.124187877	17	6	16

Table 6.2: Enrichment of Grain binding motifs in different enhancer categories
Table showing the relative enrichment of grain binding motifs in enhancers of genes that are upregulated, downregulated or not-differentially expressed between the wing and halteres.

Upregulated targets	log2FC (wing/haltere)	Downregulated targets	log2FC (wing/haltere)	Not-differential targets	log2FC (wing/haltere)
zormin	-0.789173874	Wnt2	2.423017188	sty	0.010892832
zfh2	-0.676669872	vg	0.816232399	Ppn	-0.002317595
VGAT	-0.703843376	nord	1.738986489	Pli	0.170741414
Tsp	-1.125991743	magu	0.918945096	numb	-0.027698643
Sox15	-0.96135605	ltl	3.562463451	natalisin	-0.006756337
mamo	-0.61175251	hdc	0.625056762	Mmp2	0.015197849
Dr	-0.717863205	Dll	1.2671962	jar	0.03884132
CR44955	-1.088894964	CR43431	0.879900495	emp	-0.065410784
CG5966	-0.798300795	CG9270	1.069005761	CG6055	-0.03490114
CG41520	-0.656395775	bs	0.814353829	CG43658	-0.002246487
CG34193	-0.612557739	Antp	2.084424707	CG13252	0.034472586
CG1718	-0.819116634	al	3.018954292	CG13192	-0.013075742
bab2	-1.053544189			CG12702	0.06604422
				blw	-0.017522157
				Art4	-0.04276092
				Aldh-III	0.000769071

Table 6.3: Genes which are predicted to be bound by Grain and Ubx
Table showing the list of genes that are putative targets of both grain and Ubx. Note that important wing patterning genes like vg whose repression is critical for haltere specification have binding motifs for both Ubx and grain.

6.4 Discussion

In *Drosophila*, specification of the haltere fate in T3 by Ubx entails repression of wing developmental pathways. This provides a perfect system for understanding the various aspects of development and growth control. Ubx modulates a large number of target genes involved in AP/DV patterning events, cell size and shape regulation as well as at hormonal and ECM control. Interestingly, Ubx not only activates target gene but also represses them. However, being a homeodomain containing protein which recognizes a motif containing a TAAT core site in-vitro with low specificity and affinity, the mechanisms by which Ubx recognizes which genes to upregulate and which to downregulate remain largely unresolved.

We attempted to reveal the mechanisms of target recognition by Ubx using a genome wide strategy. Our results have indicated that a motif containing a TAAAT motif is bound by Ubx in *Drosophila* halteres and is functionally involved in activation as well as repression of target genes. While this indicates that the TAAAT motif might be relevant for target selection by Ubx in *Drosophila* halteres, our studies reveal that it is not used for target recognition. We developed an in-silico pipeline wherein we scan for transcription factor binding motifs in enhancers of Ubx targets that are upregulated or downregulated in the haltere. Our results do suggest that binding motifs for transcription factors like Mes2 are overrepresented in enhancers of upregulated genes, however, the importance and implications of such results needs to be tested extensively and empirically. Even though our studies using extensive in-silico analysis help provide some clues about the milieu of cofactors which might aid Ubx in target recognition, the mechanisms underlying such phenomenon remain an investigation to be completed. While our studies do indicate that a large number of cofactors are involved in Ubx mediated regulation of target genes, the transcriptional code that is recognized by during the haltere specification process still remains largely unanswered. Perhaps, better technology involving machine learning approach could be employed to gain insights into the question of target recognition by Hox genes in the future.

SUMMARY

We carried out this study in an attempt to understand the mechanism of target selection and target recognition by the Hox protein Ultrabithorax in Drosophila halteres. Our studies provide compelling evidence about a high affinity binding motif, termed as the TAAAT motif, to be involved in target selection by Ubx. While the TAAAT motif was not required for target recognition, our results indicate that incorporation of the TAAAT motif in enhancers of important wing patterning genes might have facilitated their regulation by Ubx in the dipteran lineage, thereby aiding in the haltere specification process. Taken together, our studies provide key insights into the mechanisms of target selection by Ubx and help throw some light into how the evolution of hindwing to haltere morphology might have been influenced by the evolution of target selection mechanism in insect species.

The following are the few salient findings of this study:

The TAAAT motif is the preferred binding site for Ubx in *Drosophila* halteres

Our studies provide interesting revelations into a novel DNA recognition motif containing a TAAAT core site (GSCCATAAATHA) which is enriched by Ubx in *Drosophila* halteres. The frequency of this motif was found to be 1.7folds greater in Ubx pulled down sequences as compared to the entire genome, suggesting that the TAAAT motif acts a preferred binding motif for Ubx in *Drosophila* halteres. The fact that the frequency of the TAAAT motif was similar between enhancers of genes that are upregulated, downregulated as well as not-differentially expressed between the wing and haltere, indicates that the TAAAT motif might be required for both upregulation and downregulation of targets and thus may not be involved in target recognition by Ubx.

The TAAAT motif is critical for Ubx mediated regulation of the *CG132222* gene

Experiments using the edge enhancer of the CG13222 gene which is upregulated by Ubx in Drosophila halteres suggested that the TAAAT motif was bound with higher affinity and specificity by Ubx as compared to the TAAT motif. We also observed that this high affinity TAAAT motif is critical for the activation of the edge enhancer of the CG13222 gene in both *Drosophila* halteres as well as S2 cells. Additionally, the TAAAT motif was sufficient to bring an otherwise unresponsive binding site of Ubx (having a TAAT motif) under Ubx regulation.

The TAAAT motif, thus, is not only able to confer increased specificity of binding, but can also facilitate target gene selection which may be possible due to a more favorable chromatin structure for Ubx binding attained by the TAAAT motif as compared to the TAAT motif. Our studies, thus, reveal the importance a high affinity binding motif for the Ubx protein in *Drosophila* halteres which is functionally relevant for the activation of a target gene during haltere development.

Presence of the TAAAT motif seems to have brought the enhancer of the *vestigial* gene from *Apis* under Ubx control in the dipteran lineage

Our studies with the quadrant enhancer of the *vestigial* gene (*quad-vg*) were inconclusive in the assessment of the role of the TAAAT motif in downregulation of the *vestigial* gene since most mutations in the enhancer led to loss of reporter GFP expression from both wing and haltere imaginal discs. While this suggested that the region being mutated might be important for the regulation of the *quad-vg* enhancer, we did not test this hypothesis empirically. Sequence comparison of *quad-vg* and *Apis-vg* enhancer indicated that while the quad-vg enhancer had both TAAT and TAAAT motifs, the *Apis-vg* enhancer had a single TAAT motif. Transgenic assays using the *Apis-vg* enhancer revealed that replacing the TAAT motif with TAAAT motif, quite appreciably, brought the enhancer under the negative regulation of Ubx in Drosophila halteres. Interestingly, the repression of reporter expression driven by the mutant enhancer was observed in the pouch region only and not in the hinge region, suggesting that manipulation of the TAAT motif to the TAAAT motif of *Apis-vg* enhancer brings its expression patterns closer to that of *quad-vg* of *Drosophila*. Thus, microevolutionary changes replacing the TAAT motif with TAAAT motif in the enhancer of key wing patterning genes like *vg* might be critical to bring them under the regulation of Ubx, there facilitating haltere specification in the dipteran lineage.

Differential enrichment of the TAAAT motif in enhancers of Ubx targets in *Apis* and *Drosophila* suggest a mechanism of evolution of target selection

Genome wide comparative analysis of Ubx binding between the developing hindwing of *Apis* and developing halteres of *Drosophila* reveal that TAAAT motif are enriched in Ubx pulled down sequences in *Drosophila* but not in the *Apis*. This is suggestive of a recurrent theme wherein microevolutionary changes in the cis-regulatory regions, specifically in Ubx binding motifs which confer stronger binding, might have brought certain important wing patterning

genes like *vg* under Ubx regulation in the dipteran lineage. Since genes like *vg* control important regulatory networks required for the wing specification process, the modulation of such genes by Ubx would in turn lead to changes in the chromatin landscape of many other targets. This would in turn make them sensitive to Ubx mediated regulation, thereby facilitating the haltere specification process.

An analysis pipeline to reveal mechanism of target selection by Ubx

We developed an in-silico pipeline wherein we scan for transcription factor binding motifs in enhancers of Ubx targets that are upregulated or downregulated in the haltere. Our results do suggest that binding motifs for transcription factors like Mes2 are overrepresented in enhancers of upregulated genes, however, the importance and implications of such results needs to be tested extensively and empirically. Even though our studies using extensive in-silico analysis help provide some clues about the milieu of cofactors which might aid Ubx in target recognition, we have not been able to exact mechanisms underlying such phenomenon. Perhaps, better technology involving machine learning approach will be able to provide clues into the question of target recognition by Hox genes in the future.

FUTURE DIRECTIONS

1. One of the most interesting questions relating to Hox protein function is that how do Hox proteins like Ubx, which do not have a dedicated protein interaction domain, differentiate between which targets to upregulate and downregulate. While we tried to address this question in this study, our technology was unable to untangle the vast intricacies of gene regulatory topography which are recognized by Ubx. A better understanding of enhancers of Ubx targets using methods like ATAC-seq, Chromatin confirmation capture (3C) and Hi-C combined with Machine learning algorithms will provide a better understanding of the question at hand. This would not only help in understanding Hox gene mediated recognition of targets but will also be critical for understanding organ growth and development at the molecular level.

2. Our data suggests that a large number of regions bound by Ubx belong to genes that are not differentially expressed between the wing and haltere. In this context, we found an interesting pattern on comparing the genome wide targets of Ubx in the third instar, prepupal and pupal stages. We found that the not differentially expressed genes whose enhancers are bound by Ubx in the third instar imaginal discs, are differentially expressed in the prepupal and pupal stages. This suggests, to some extent, that while Ubx binds to a large number of target enhancers, their regulation is dependent on temporal cues. A fascinating aspect of Ubx function is the fact that it is expressed at all times during Drosophila development starting from the embryonic level. Perhaps, one of the mechanisms of target selection and regulation by Ubx may involve binding loosely to all recognition sites and regulating the target depending on temporal cues.

3. Our studies suggest that evolution of Ubx binding motif preference might be one of the key determinants for the evolution of halteres in the dipteran lineage. However, morphological diversification of hindwing structures in insects might have been a culmination of a large number of factors. One of the interesting differences between Ubx binding in Apis and Drosophila was the divergence in topology of Ubx binding. In Drosophila, Ubx bound away from TSS of targets and in Apis, Ubx binding sites were closer to the TSS. The implications of such differences in binding topology on the differential regulation of target genes in evolution remain largely unexplored and might open up new avenues for understanding molecular evolution.

Materials and Methods

1. ChIP sequencing

a) **Sample collection:** For a particular biological replicate, 30 bottles containing 6% corn-agar media were seeded with 30 adult Drosophila (CS strain) and incubated at 25 degrees. Flies were discarded after 3 days and further incubated till larvae emerged. Third instar wandering larvae were collected, cut, inverted and fixed with cold 1.5% PFA solution. The solution was changed after 10minutes and incubated further for 10minutes. 50ul of 1.375M Glycine solution (final concentration 125mM) was added to quench the reaction and incubated for 5minutes. The samples were washed twice with 1XPBS and then dissected for wing and haltere discs. Samples were collected in separate tubes containing cold PBS, snap frozen and store at -80 degrees for further processing.

b) **Homogenization, Sonication and Immunoprecipitation:** On the day of processing, samples were thawed on ice and PBS removed. 500ul of Cell Lysis Buffer was added to each tube and followed by mechanical shearing using a polyester pestle while keeping the samples on ice. Samples were spun at 2000rpm for 10mins at 4 degrees and supernatant discarded. The pellet was resuspended in 400ul of Sonication buffer and incubated on ice for 30mins. The solution was transferred to a cuvette and sonication carried out using the Covaris S2 sonicator (DC 20%, Intensity 5, Cycles of Burst 200, Time= 40*30 (20mins)). Samples were transferred to a 1.5ml tube and spun at 14000rpm for 15mins at 4 degrees. The supernatant was transferred to a clean microcentrifuge tube and 10% of the solution set aside as Input. The entire solution was divided into two parts (one Samples were precleared by incubating them with 4ul of Magnetic A Beads with rotation at 4 degrees. The supernatant was transferred to a clean microcentrifuge tube and 2ul of Polyclonal Ubx antibody added to the solution. Samples were incubated overnight at 4 degrees and 15ul of magnetic beads added with incubation for 4hrs with gentle rotation. The supernatant was removed and beads washed with buffers in the following sequence- Low salt buffer (2 times), High Salt buffer (2 times), LiCl buffer (2 times), TE buffer (2 times). 150ul of elution buffer was then added to the beads and heated at 65 degrees for 15mins. The step was repeated and

the supernatant collected in a clean tube. For decrosslinking, NaCl solution at a final conc of 210mM was added and incubated overnight at 65 degrees. Samples were further treated with 1ul RNAse (incubated at 37 for 1hour) followed by Proteinase K treatment for 2hours at 42 degrees. Samples were purified using PCI purification and quantified using the Qubit HS DNA quantification system.

c) **Library Preparation:** Equal amount of DNA (~2 ng) was used as an input for library preparation using NEB Ultra II DNA library prep kits (NEB #E7645). Number of cycles for amplification of adapter ligated libraries were estimated by the qPCR before final amplification to avoid any bias arising due to PCR amplification and indexing (NEB #E7350). Final amplified libraries were purified twice, first with 1X followed by 0.8x volume of beads per sample using HiPrep PCR clean up system (Magbio #AC-60050). Library concentration was determined using Qubit HS DNA kit (Invitrogen #E7350) and average fragment size was estimated using DNA HS assay on bioanalyzer 2100 (Agilent #5067-4626) before pooling libraries in equimolar ratio. Sequencing reads (100bp PE) were obtained on the Hiseq 2500 V4 platform at Macrogen Inc, Korea.

d) **Analysis:** Raw reads were trimmed to remove adapter sequences using Trimmomatic and aligned to the Drosophila genome (BDGP6.28) using the bwa-mem software. Sam files were converted to BAM format and duplicate reads removed using Bamtools. Peak calling was performed using the MACS2 software with an FDR of 0.01 and only those peaks which were present in at-least two replicates were considered for further processing. Peaks were annotated using the annotatePeaks.pl program from the Homer package and the Deeptools package was used to generate bigwig files with the criteria bs1 and smooth length 150.

2. RNA sequencing

a) **Sample collection and Sequencing:** For a particular biological replicate, 20 bottles containing 6% corn-agar media were seeded with 30 adult Drosophila (CS strain) and incubated at 25 degrees. Flies were discarded after 3 days and further incubated till larvae emerged. Third instar wandering larvae were collected, cut, inverted and dissected for wing and haltere discs in cold PBS. Samples were collected in separate

tubes containing Trizol solution, snap frozen and sent to Genotypic Bangalore for further processing. Sequencing was performed on the Illumina Hiseq2000 platform.

b) **Analysis:** Raw fastq files were aligned to the dm6 genome using HiSAT2 software after sequencing quality check performed using the FASTQC software. RNA extraction, library preparation and sequencing read generation on the NextSeq 75bp paired end platform were carried out at Genotypic Technologies, Bangalore. Hisat2 software was used to align reads to the dm6 genome, sorted using the samtools software and read counting done using htseq software. The edgeR software was used to identify differentially expressed genes between the wing and haltere using a cutoff of minimum one read for each gene per replicate.

3. Motif analyses and protein sequence comparison: De-novo motif analysis was carried out using Homer (findMotifsGenome.pl). The PWM for the TAAAT motif obtained from Homer was converted to Transfac format using RSAT (Nguyen et al. 2018) and finally converted to MEME format using transfac2meme command. A 100bp pad on both sides was applied to ChIP peak summit for calculation of frequency of motifs using the FIMO software from MEME suite. For calculating the frequency of motifs in the entire Drosophila or Apis genome, random sequences of 200bp were generated using the bedtools random software. The number of such sequences was equal to the total number of ChIP peaks being used for motif analysis from Drosophila or Apis datasets. A total of 100 iterations were performed using the mentioned parameters and the average frequency of occurrence was considered.

4. Molecular Cloning: The list of primers used for generation of various mutants and the parent vectors in which they were cloned has been provided as a table in Appendix 1. For cloning the wild type enhancers of CG13222, quadVg and ApisVg, restriction cloning strategy was used. Specific primers containing Nhe1 and Kpn1 restriction sites were designed against the region of interest and PCR amplification performed. PCR products were purified using the Qiagen PCR purification system and quantified using Nanodrop. 2ug of PCR products were digested with Kpn1 and Nhe1 enzymes. Simultaneously, the pGL3-DE5-GFP vector was digested and gel purified using the Qiagen Gel purification system followed by rSAP treatment. Ligation reaction was setup at vector to insert ratio of 1:6 using the Promega rapid ligation system. Appropriate controls were setup and 2ul of each reaction was transformed into

ultracompetent DH5-alpha cells and further plated onto LB-agar plates containing Ampicillin as the screening antibiotic. Colony PCR was performed to screen for positive clones and further sequence verified using Sanger sequencing.

All mutant constructs (for CG13222, quadVg and ApisVg) were first cloned into pGL3-DE5 vectors side directed mutagenesis. Specific primers were designed carrying the mutation of interest and PCR amplification performed using the GL2 universal primer (against pGL3 vector) as the reverse primer. PCR purification was further carried out and quantification done using the Nanodrop system. A second round of PCR was performed using appropriate molar ratios of the PCR product and template plasmid (WT). A control reaction containing everything except the PCR product was setup simultaneously and both reactions subject to Dpn1 digestion. Samples were transformed into ultra-competent Dh5 alpha cells and plated on LB agar plates using Ampicillin as the selection medium. Plasmid isolation was carried out for isolated colonies and Sanger sequencing performed for sequence verification.

For Luciferase assays, verified mutant constructs were sub-cloned between Kpn1 and Nhe1 restriction sites, upstream of a modified pGL3 vector containing a 5X Dorsal binding site followed by sequence verification

For generating transgenics, all verified constructs were cloned between the Kpn1 and Nhe1 sites of pH-stinger-attb (a kind gift from Manfred Frasch) containing a GFP reporter. Metallothionein inducible pRMHa3 vectors containing Ubx were previously generated in the lab and empty pRMHa3 vector was generated by excising the cloned Ubx sequence (with Garima).

5. List of primers and constructs generated:

a. Primer sequences used for cloning CG13222 constructs into pGL3-DE5 or pH-Stinger-attb vectors

CG13222_FP: ACGGTACCCATAGACCACCAGCCACTGT

CG13222_RP: ATCGCTAGCTCAATCGCGTACCGAAGCAA

b. Primer sequences used for cloning quad-vg constructs into pH-Stinger-attb vectors

quad-vg_FP: ACGGTACCGGAGCTCCCTCCGGAGAC

quad-vg_RP: ATCGCTAGCCGATTGTACTTTGTCGTTTCTAA

c. Primer sequences used for cloning Apis-vg constructs into pH-Stinger-attb vectors

Apis-vg_FP: ACGGTACCCTTCTCGCGAGAAACGAGAGGC

Apis-vg_RP: ATCGCTAGCGTGGACAGTGACGAGGACACG

d. Reverse Primer (for all constructs):

GL primer2 (CTTTATGTTTTTGGCGTCTTCCA)

e. Description of constructs generated and specific primers used for each

Name of construct	**Description**	**Vector backbone**	**Primer sequence (Forward)**
***CG13222* enhancer constructs (For Luciferase Assay and EMSA)**			
M1_A	Only TAAAT motif mutated at site1	pGl3-DE5	GCTTGTTAACACGCAG AGGGATTACACTGGCC GCCCGCGAGATT
M1_B (only used for EMSA)	Only TAAT motif mutated at site1. This was used for EMSA only		CGCAGATAAATGGCAC TGGCC
M2_A	TAAAT motif introduced in place of TAAT motif in site2	pGl3-DE5	CACTGGCCGCCCGCGA GATTTACCATCGAGAT GCAGTCAG
***CG13222* enhancer constructs (For *Drosophila* Transgenics)**			
M1	Both TAAT and TAAAT motifs mutated (Hersh et. al 2007)	pH-Stinger-attb	GCTTGTTAACACGCAG ATAACGCGCACTGGCC GCCCGCGAGATT
M1_A	Only TAAAT motif mutated at site1	pH-Stinger-attb	GCTTGTTAACACGCAG AGGGATTACACTGGCC GCCCGCGAGATT
M2_A	TAAAT motif introduced in place of TAAT motif in site2	pH-Stinger-attb	CACTGGCCGCCCGCGA GATTTACCATCGAGAT GCAGTCAG

***quad-vg* enhancer constructs**			
*quad-vg*_M	TAAAT motif mutated	pH-Stinger-attb	GATCGAACGTTATCGGGTCCGCGCCGCCACGCTCTCTTCATTAG
*quad-vg*_M1	TAAAT motif mutated to TAAT motif	pH-Stinger-attb	GATCGAACGTTATCGGGTCTAATCGCCACGCTCTCTTCATTAG
*quad-vg*_M2	TAAT motif mutated	pH-Stinger-attb	TCGCCACGCTCTCTTCGCGCGGCCAAAAGGTGAAAG
***Apis-vg* enhancer constructs**			
*Apis-vg*_M1	Mutation replacing the entire TAAT motif with TAAAT motif	pH-Stinger-attb	TCGGCCCCCATTAAGCTCTTCGATTTATGAGCGAGCATCTGAGGGGCCGA
*Apis-vg*_M2	Mutation replacing the TAAT motif with a 25bp cassette from *quad-vg*	pH-Stinger-attb	TCGGCCCCCATTAAGCTCTTGAGAGCGTGGCGATTTATGAGCGAGCATCTGAGGGGCCGA
*Apis-vg*_M3	Mutation replacing the TAAT motif with par of the cassette from quad-vg (TAAAT motif present)	pH-Stinger-attb	TCGGCCCCCATTAAGCTCTTTAATGAAGAGAGCGTGGCGATTTATGAGCGAGCATCTGAGGGGCCGA
*Apis-vg*_M4	Mutation replacing the TAAT motif with part of the cassette from quad-vg (TAAAT motif absent)	pH-Stinger-attb	TCGGCCCCCATTAAGCTCTTTAATGAAGAGAGGCGAGCATCTGAGGGGCCGA

6. Luciferase Assays: Schneider's cells (S2) were plated onto 24 well plates at a density of 3*10^5 cells per well 6 hours prior to transfection. For every construct, either wild type or mutant, two sets of experiments were designed; one well was co-transfected with the enhancer construct in pGL3 vector and the empty pRMHa3 vector whereas the other well was co-transfected with the enhancer construct and pRMHa3 vector containing Ubx. Renilla luciferase was used as an internal control and co-transfected in all experiments. Transfection was carried out using the Effectene transfection reagent and all experiments carried out in 3 technical replicates and at least 3 biological replicates. 48 hours post transfection, sterile CuSO4 solution was used to induce expression of Ubx at a final concentration of 500um and incubated for 24 hours. Cells were harvested, pelleted down (1000rpm for 4mins) and 100ul of 1X Passive Lysis

buffer added and vortexed to dissolve the cell pellet. Cells were incubated for 15mins at room temperature and further spun at 10,000rpm for 90secs to collect the supernatant. The luminescence was measured with the Dual glo Luciferase assay kit (Promega) on Ensight Plate reader (Perkin Elmer). All readings were normalized to Renilla luminescence and datasets compared using the GraphPad Prism software.

7. Immunohistochemistry: Wandering third instar larvae were cut and inverted in cold PBS followed by fixation with 4% PFA for 20mins with gentle rocking. Samples were washed thrice with 0.1% PBTX (0.1% TritonX in PBS) for 10 minutes each, followed by one hour of blocking at room temperature (blocking solution: 0.5% BSA in 0.1%PBTX). Samples were incubated with primary antibodies (dilutions made in blocking solution) at 4 degree overnight followed by washing with 0.1% PBTX for 10mins (x3 times) at room temperature. Samples were incubated with secondary antibodies for one hour at room temperature and then washed with 0.1% PBTX for 10mins (x3 times) followed by PBS was (x3 times). Wing and haltere imaginal discs were dissected out and mounted using Prolong Gold Antifade (Invitrogen) and stored at 4 degrees. All imaging was done using Leica Sp8 system using 1.5 and 3 magnifications on 40x oil objective. Antibodies used in the study are Rb-GFP (1:1000), Rb-Ubx (1:1000), m-Ubx (1:30), goat-Hth (1:50), chicken-GFP (1:500), alexa-fluor 568 (1:1000) and alexa-fluor 633 (1:1000).

8. GUI based analysis pipeline: Comparison of Ubx direct targets with the list of differentially expressed genes provided us with enhancers (Ubx peaks) which were categorized as upregulated, downregulated or not differentially expressed. This is used as the input for the program. First, the information is extracted using bcd.py followed by FASTA generation of sequences for the given input using bedtools software. The ‘slop’ and ‘getfasta’ commands are used to generate an “up_final.bed” file for upregulated sequences. The same is done for the other two categories; downregulated and not-differentially expressed. FASTA files are then split using our python script FASTA_splitting.py, whose output is stored as 1.txt, 2.txt and so on. Motif occurrence calculation for every splitted FASTA file is done using FIMO (Find Individual Motif Occurrences) from MEME suite and the output is stored in a ‘. tsv’ file, which stores the number of occurences of different motifs (obtained from JASPAR database) present in the sequence. To maintain the custom order of motifs post scanning, a sorting script is run;

Sort_seq.py. Next, using a combination of scripts, the frequency of each transcription factor in upregulated, downregulated and not-differential categories are provided as a tabulated format. The scripts written for these compute expensive processes are highly optimised leveraging the concepts of multi-threading and core optimisation. We have taken leverage of GNU Parallel in order to parallely run certain independent python scripts, thus improving the time complexity by 1.5 times (*for i7-7700 HQ Processor). To provide deeper insights about the result, we generate advanced visual plots which help save time further. This is accomplished using the Seaborn package. We also provide a predictive module trained on our dataset, which gives an approximate idea of the regulation characteristic of a motif. It can be trained on an external dataset given it has the input format as specified.

Detailed information about this pipeline can be obtained from https://github.com/abhikbhattacharjee/Motifizer

9. List of Bioinformatic Commands

a. For trimming

TrimmomaticPE haltere_input_rep1_1.fastq haltere_input_rep1_2.fastq
./trimmed_chip_seq_reads/Trimmed/haltere_input_rep1_1_trimmed.fastq
./trimmed_chip_seq_reads/Untrimmed/haltere_input_rep1_1_untrimmed.fastq
./trimmed_chip_seq_reads/Trimmed/haltere_input_rep1_2_trimmed.fastq
./trimmed_chip_seq_reads/Untrimmed/haltere_input_rep1_2_untrimmed.fastq
ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:20 TRAILING:20 MINLEN:30

b. Alignment to the genome

bwa mem
./Drosophila_melanogaster_UCSC_dm6/Drosophila_melanogaster/UCSC/dm6/Sequence/BWAIndex/genome.fa
./trimmed_chip_seq_reads/Trimmed/haltere_input_rep1_1_trimmed.fastq
./trimmed_chip_seq_reads/Trimmed/haltere_input_rep1_2_trimmed.fastq >
./bwa_aligned/input_rep1_haltere_bwaalign

c. Sam to bam

samtools view -S -b ./bwa_aligned/%s > ./bwa_aligned/sam_to_bam/%s.bam

d. Filtering

samtools view -F 0x200 -F 0x4 -b ./bwa_aligned/sam_to_bam/%s > ./bwa_aligned/filtered_bam/filtered_%s

e. Peak calling (MACS2)

macs2 callpeak -t ./bwa_aligned/filtered_bam/filtered_Ubx_rep1_haltere_bwaalign.bam -c ./bwa_aligned/filtered_bam/filtered_input_rep1_haltere_bwaalign.bam -f BAMPE -g dm -n Halt_Rep1 --outdir ./bwa_aligned/Peak_calling_dir/Macs2/MACS2_0.01_dup1 -B -q 0.01

f. PEPR

./bwa_aligned/Peak_calling_dir/PePr_param1")
#os.system ("PePr -c ./bwa_aligned/filtered_bam/filtered_Ubx_rep1_haltere_bwaalign.bam,./bwa_aligned/filtered_bam/filtered_Ubx_rep2_haltere_bwaalign.bam,./bwa_aligned/filtered_bam/filtered_Ubx_rep3_haltere_bwaalign.bam -i ./bwa_aligned/filtered_bam/filtered_input_rep1_haltere_bwaalign.bam,./bwa_aligned/filtered_bam/filtered_input_rep2_haltere_bwaalign.bam,./bwa_aligned/filtered_bam/filtered_input_rep3_haltere_bwaalign.bam --keep-max-dup 2 -f bampe --threshold 1e-4 --peaktype sharp --name Halt_PePr_narrow --output-directory ./bwa_aligned/Peak_calling_dir/PePr_param1

g. Homer

annotatePeaks.pl dmel_peaks_bed '/media/iiser/Naveen/Naveen/chip_seq/haltere_chipseq/Drosophila_melanogaster.BDGP6.28.dna_rm.toplevel.fa' -gtf

'/media/iiser/Naveen/Naveen/chip_seq/haltere_chipseq/Drosophila_melanogaster.BDGP6.28.100.gtf' > Dmel_fimo_homer

h. Getfasta

bedtools getfasta -fi '/media/iiser/Naveen/Naveen/chip_seq/Apis_chipseq/Apis_4.5/bwa_idx/Apis_mellifera.Amel_4.5.dna_rm.toplevel.fa' -bed Amel_rand -fo Amel_rand_fasta

i. FIMO

findMotifsGenome.pl '/media/iiser/Naveen/Naveen/chip_seq/Dmel_Apis_common_peaks/commontargets_Dmelpeaks_100bpbothsides_bed' dm6 '/media/iiser/Naveen/Naveen/chip_seq/Dmel_Apis_common_peaks/commontargets_Dmelpeaks_100bpbothsides_motif' -size 200 -mset insects

References

Agrawal P, Habib F, Yelagandula R, Shashidhara LS. 2011. Genome-level identification of targets of Hox protein Ultrabithorax in Drosophila: novel mechanisms for target selection. *Sci Rep* **1**:1–10. doi:10.1038/srep00205

Appel B, Sakonju S. 1993. Cell-type-specific mechanisms of transcriptional repression by the homeotic gene products UBX and ABD-A in Drosophila embryos. *EMBO J* **12**:1099–1109.

Aza-Blanc P, Kornberg TB. 1999. Ci: a complex transducer of the hedgehog signal. *Trends Genet* **15**:458–462. doi:10.1016/s0168-9525(99)01869-7

Bender W, Akam M, Karch F, Beachy PA, Peifer M, Spierer P, Lewis EB, Hogness DS. 2007. Molecular genetics of the bithorax complex in Drosophila melanogaster. *Genes, Dev Cancer Life Work Edward B Lewis* **2**:311–326. doi:10.1007/978-1-4020-6345-9_14

Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Peña-Castillo L, Alleyne TM, Mnaimneh S, Botvinnik OB, Chan ET, Khalid F, Zhang W, Newburger D, Jaeger SA, Morris QD, Bulyk ML, Hughes TR. 2008. Variation in Homeodomain DNA Binding Revealed by High-Resolution Analysis of Sequence Preferences. *Cell* **133**:1266–1276. doi:10.1016/j.cell.2008.05.024

Brook WJ, Diaz-Benjumea FJ, Cohen SM. 1996. Organizing Spatial Pattern in Limb Development. *Annu Rev Cell Dev Biol* **12**:161–180. doi:10.1146/annurev.cellbio.12.1.161

Cabrera C V., Botas J, Garcia-Bellido A. 1985. Distribution of Ultrabithorax proteins in mutants of Drosophila bithorax complex and its transregulatory genes. *Nature* **318**:569–571. doi:10.1038/318569a0

Carroll SB. 1995. Homeotic genes and the evolution of arthropods and chordates. *Nature* **376**:479–485. doi:10.1038/376479a0

Carroll SB, Weatherbee SD, Langeland JA. 1995. Homeotic genes and the regulation and evolution of insect wing number. *Nature* **375**:58–61. doi:10.1038/375058a0

Carroll Sean B. , Grenier Jennifer K. WSD. 2004. From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design.

Castelli-Gair JE, Micol JL, Garcia-Bellido A. 1990. Transvection in the Drosophila Ultrabithorax gene: A Cbx1 mutant allele induces ectopic expression of a normal allele in trans. *Genetics* **126**:177–184. doi:10.1093/genetics/126.1.177

Choo SW, White R, Russell S. 2011. Genome-wide analysis of the binding of the hox protein ultrabithorax and the hox cofactor homothorax in Drosophila. *PLoS One* **6**:25–27. doi:10.1371/journal.pone.0014778

Crickmore MA, Mann RS. 2007. Hox control of morphogen mobility and organ development through regulation of glypican expression. *Development* **134**:327–334. doi:10.1242/dev.02737

Crickmore MA, Mann RS. 2006. Hox control of organ size by regulation of morphogen production and mobility. *Science (80-)* **313**:63–68. doi:10.1126/science.1128650

Crocker J, Abe N, Rinaldi L, McGregor AP, Frankel N, Wang S, Alsawadi A, Valenti P, Plaza S, Payre F, Mann RS, Stern DL. 2015. Low Affinity Binding Site Clusters Confer Hox Specificity and Regulatory Robustness. *Cell* **160**:191–203. doi:https://doi.org/10.1016/j.cell.2014.11.041

Das PM, Ramachandran K, Singal R. 2004. Das376.indd **37**:1–9.

Davidson EH. 2001. Genomic Regulatory Systems. doi:https://doi.org/10.1016/B978-0-12-205351-1.X5000-9

De Las Heras JM, García-Cortés C, Foronda D, Pastor-Pareja JC, Shashidhara LS, Sánchez-Herrero E. 2018. The Drosophila Hox gene Ultrabithorax controls appendage shape by regulating extracellular matrix dynamics. *Dev* **145**:1–10. doi:10.1242/dev.161844

De Navas LF, Reed H, Akam M, Barrio R, Alonso CR, Sánchez-Herrero E. 2011. Integration of RNA processing and expression level control modulates the function of the Drosophila Hox gene Ultrabithorax during adult development. *Development* **138**:107–116. doi:10.1242/dev.051409

Diaz-Benjumea FJ, Cohen SM. 1993. Interaction between dorsal and ventral cells in the imaginal disc directs wing development in Drosophila. *Cell* **75**:741–752. doi:10.1016/0092-8674(93)90494-b

Ekker SC, Young KE, Von Kessler DP, Beachy PA. 1991. Optimal DNA sequence recognition by the Ultrabithorax homeodomain of Drosophila. *EMBO J* **10**:1179–1186. doi:10.1002/j.1460-2075.1991.tb08058.x

Francois J. 1977. Evolution and Tinkering **196**:1161–1166.

Galant R, Walsh CM, Carroll SB. 2002. Hox repression of a target gene: extradenticle-independent, additive action through multiple monomer binding sites. *Development* **129**:3115–3126.

Hersh BM, Carroll SB. 2005. Direct regulation of knot gene expression by Ultrabithorax and

the evolution of cis-regulatory elements in Drosophila. *Development* **132**:1567–1577. doi:10.1242/dev.01737

Hersh BM, Nelson CE, Stoll SJ, Norton JE, Albert TJ, Carroll SB. 2007. The UBX-regulated network in the haltere imaginal disc of D . melanogaster **302**:717–727. doi:10.1016/j.ydbio.2006.11.011

Hughes CL, Kaufman TC. 2002. Hox genes and the evolution of the arthropod body plan. *Evol Dev* **4**:459–499. doi:10.1046/j.1525-142x.2002.02034.x

Irvine KD, Harvey KF. 2015. Control of organ growth by patterning and hippo signaling in drosophila. *Cold Spring Harb Perspect Biol* **7**:1–16. doi:10.1101/cshperspect.a019224

Johnston DS, Nüsslein-Volhard C. 1992. The origin of pattern and polarity in the Drosophila embryo. *Cell* **68**:201–219. doi:10.1016/0092-8674(92)90466-P

Kim J, Sebring A, Esch JJ, Kraus ME, Vorwerk K, Magee J, Carroll SB. 1996. Integration of positional signals and regulation of wing formation and identity by Drosophila vestigial gene. *Nature* **382**:133–138. doi:10.1038/382133a0

Lappin TRJ, Grier DG, Thompson A, Halliday HL. 2006. HOX genes: Seductive science, mysterious mechanisms. *Ulster Med J* **75**:23–31.

Lewis EB. 1978. A gene complex controlling segmentation in Drosophila. *Genes, Dev Cancer Life Work Edward B Lewis* 229–242. doi:10.1007/978-1-4020-6345-9_10

Makhijani K, Kalyani C, Srividya T, Shashidhara LS. 2007. Modulation of Decapentaplegic gradient during haltere specification in Drosophila. *Dev Biol* **302**:243–255. doi:10.1016/j.ydbio.2006.09.029

Mann RS, Affolter M. 1998. Hox proteins meet more partners. *Curr Opin Genet Dev* **8**:423–429. doi:10.1016/s0959-437x(98)80113-5

McGinnis W, Krumlauf R. 1992. Homeobox genes and axial patterning. *Cell* **68**:283–302. doi:10.1016/0092-8674(92)90471-N

Merabet S, Galliot B. 2015. The TALE face of Hox proteins in animal evolution. *Front Genet* **6**:1–13. doi:10.3389/fgene.2015.00267

Mohit P, Makhijani K, Madhavi MB, Bharathi V, Lal A, Sirdesai G, Reddy VR, Ramesh P, Kannan R, Dhawan J, Shashidhara LS. 2006. Modulation of AP and DV signaling pathways by the homeotic gene Ultrabithorax during haltere development in Drosophila **291**:356–367. doi:10.1016/j.ydbio.2005.12.022

Morris SC. 1993. The fossil record and the early evolution of the Metazoa. *Nature* **361**:219–225. doi:10.1038/361219a0

Noyes MB, Christensen RG, Wakabayashi A, Stormo GD, Brodsky MH, Wolfe SA. 2008. Analysis of Homeodomain Specificities Allows the Family-wide Prediction of Preferred Recognition Sites. *Cell* **133**:1277–1289. doi:10.1016/j.cell.2008.05.023

Ohno S. 1970. Evolution by Gene Duplication. doi:10.1007/978-3-642-86659-3

Pallavi SK, Shashidhara LS. 2003. Egfr/Ras pathway mediates interactions between peripodial and disc proper cells in Drosophila wing discs. *Development* **130**:4931–4941. doi:10.1242/dev.00719

Passner JM, Ryoo HD, Shen L, Mann RS, Aggarwal AK. 1999. Structure of a DNA-bound Ultrabithorax ± Extradenticle homeodomain complex. *Nature* **397**:714–719.

Pearson JC, Lemons D, McGinnis W. 2005. Modulating Hox gene functions during animal body patterning. *Nat Rev Genet* **6**:893–904. doi:10.1038/nrg1726

Prasad N, Tarikere S, Khanale D, Habib F, Shashidhara LS. 2016a. A comparative genomic analysis of targets of Hox protein Ultrabithorax amongst distant insect species. *Sci Rep* **6**:27885. doi:10.1038/srep27885

Prasad N, Tarikere S, Khanale D, Habib F, Shashidhara LS. 2016b. A comparative genomic analysis of targets of Hox protein Ultrabithorax amongst distant insect species. *Sci Rep* **6**:27885. doi:10.1038/srep27885

Prince VE. 2002. The Hox paradox: More complex(es) than imagined. *Dev Biol* **249**:1–15. doi:10.1006/dbio.2002.0745

Roch F, Akam M. 2000. Ultrabithorax and the control of cell morphology in Drosophila halteres. *Development* **127**:97–107.

Rogulja D, Irvine KD. 2005. Regulation of cell proliferation by a morphogen gradient. *Cell* **123**:449–461. doi:10.1016/j.cell.2005.08.030

Ruiz-Losada M, Blom-Dahl D, Córdoba S, Estella C. 2018. Specification and patterning of Drosophila appendages. *J Dev Biol* **6**. doi:10.3390/jdb6030017

Shashidhara LS, Agrawal N, Bajpai R, Bharathi V, Sinha P. 1999. Negative regulation of dorsoventral signaling by the homeotic gene Ultrabithorax during haltere development in Drosophila. *Dev Biol* **212**:491–502. doi:10.1006/dbio.1999.9341

Singh S, Sánchez-Herrero E, Shashidhara LS. 2015. Critical role for Fat/Hippo and IIS/Akt pathways downstream of Ultrabithorax during haltere specification in Drosophila. *Mech Dev* **138**:198–209. doi:10.1016/j.mod.2015.07.017

Slattery M, Ma L, Négre N, White KP, Mann RS. 2011a. Genome-wide tissue-specific occupancy of the hox protein ultrabithorax and hox cofactor homothorax in Drosophila.

PLoS One **6**. doi:10.1371/journal.pone.0014686

Slattery M, Riley T, Liu P, Abe N, Gomez-Alcala P, Dror I, Zhou T, Rohs R, Honig B, Bussemaker HJ, Mann RS. 2011b. Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* **147**:1270–1282. doi:10.1016/j.cell.2011.10.053

Struhl G. 1981. A homoeotic mutation transforming leg to antenna in Drosophila. *Nature* **292**:635–638. doi:10.1038/292635a0

Tomoyasu Y, Wheeler SR, Denell RE. 2005. Ultrabithorax is required for membranous wing identity in the beetle Tribolium castaneum. *Nature* **433**:643–647. doi:10.1038/nature03272

Walsh Christopher M., Carroll SB. 2007. Collaboration between Smads and a Hox protein in target gene repression. *Development* **134**:3585–3592. doi:10.1242/dev.009522

Walsh C. M., Carroll SB. 2007. Collaboration between Smads and a Hox protein in target gene repression. *Development* **134**:3585–3592. doi:10.1242/dev.009522

Weatherbee SD, Halder G, Kim J, Hudson A, Carroll S. 1998. Ultrabithorax regulates genes at several levels of the wing-patterning hierarchy to shape the development of the Drosophila haltere. *Genes Dev* **12**:1474–1482. doi:10.1101/gad.12.10.1474

Weatherbee SD, Nijhout HF, Grunert LW, Halder G, Galant R, Selegue J, Carroll S. 1999. Uitrabithorax function in butterfly wings and the evolution of insect wing patterns. *Curr Biol* **9**:109–115. doi:10.1016/S0960-9822(99)80064-5

White RAH, Akam ME. 1985. Contrabithorax mutations cause inappropriate expression of Ultrabithorax products in Drosophila. *Nature* **318**:567–569. doi:10.1038/318567a0

White RAH, Wilcox M. 1985. Regulation of the distribution of Ultrabithorax proteins in Drosophila. *Nature* **318**:563–567. doi:10.1038/318563a0

Appendix1

1. PWM of Ubx motif containing a TAAAT core (TAAAT motif) in MEME format

MEME version 4

ALPHABET= ACGT

strands: + -

Background letter frequencies (from uniform background):
A 0.25000 C 0.25000 G 0.25000 T 0.25000

MOTIF denovo_Ubx TAAAT_motif

letter-probability matrix: alength= 4 w= 10 nsites= 100 E= 0

0.192807	0.470529	0.123876	0.212787
0.171000	0.654000	0.092000	0.083000
0.882000	0.008000	0.089000	0.021000
0.013000	0.022000	0.028000	0.937000
0.807000	0.118000	0.015000	0.060000
0.793000	0.056000	0.004000	0.147000
0.781000	0.122000	0.050000	0.047000
0.115000	0.097000	0.069000	0.719000
0.248000	0.394000	0.093000	0.265000
0.653000	0.071000	0.112000	0.164000

2. PWM of canonical Ubx motif containing a TAAT core (TAAT motif) in MEME format

MOTIF MA0094.2 Ubx

letter-probability matrix: alength= 4 w= 8 nsites= 20 E= 0

0.150000	0.250000	0.150000	0.450000
0.000000	0.000000	0.000000	1.000000
0.000000	0.000000	0.000000	1.000000
0.850000	0.000000	0.000000	0.150000
1.000000	0.000000	0.000000	0.000000
0.000000	0.000000	0.000000	1.000000
0.000000	0.000000	0.300000	0.700000
0.700000	0.000000	0.300000	0.000000

URL http://jaspar.genereg.net/matrix/MA0094.2