A micro-evolutionary change in target binding sites as key determinant of Ultrabithorax function in *Drosophila*.

Soumen Khan[1#], Saurabh J. Pradhan[1], Guillaume Giraud[2], Françoise Bleicher[2], Rachel Paul[2], Samir Merabet[2], LS Shashidhara[1,3#]

[1] Indian Institute of Science Education and Research (IISER), 411008, Pune, India

[2] IGFL, ENS Lyon, UMR5242, 32 Av. Tony Garnier, 69007 Lyon, France

[3] Ashoka University, Sonipat, 131029, Haryana, India

Correspondence: ls.shashidhara@iiserpune.ac.in

soumen.khan@students.iiserpune.ac.in

Keywords: Ultrabithorax, ChIP-seq, high affinity binding sites, molecular evolution

Abstract

All Hox proteins are known to recognize, in vitro, similar DNA-binding sites containing a TAAT core sequence. This poor DNA-binding specificity is in sharp contrast with their specific functions in vivo. Here we report a new binding motif with TAAAT core sequence to which the Hox protein Ultrabithorax (Ubx) binds with higher affinity and specificity. Using transgenic and luciferase assays, we show that this new motif is critical for Ubx-mediated regulation of a target gene in *Drosophila melanogaster*. Interestingly, this new motif with TAAAT core sequences is not associated with the targets of Ubx in the honeybee, *Apis mellifera*, wherein hindwings are nearly identical to the forewings. We show that introduction of TAAAT motif in the place of TAAT motif is sufficient to bring an enhancer of a wing-promoting gene of *A. mellifera* under the regulation of Ubx. Our results, thus, suggest that binding motifs with a TAAAT core sequence may help identify functionally relevant direct targets of Ubx in *D. melanogaster* and the emergence of these binding sites may be crucial for Hox-mediated morphological changes during insect evolution.

Introduction

The diversification of morphology and emergent function of appendages have been important features of animal evolution. This is best exemplified in the diversity that is seen in closely related species of Arthropoda. In Bilaterans, *Hox* genes regulate the development of serially homologous structures along the anterior posterior axis (Lewis 1978; reviewed in

33 McGinnis and Krumlauf 1992). While genetic studies unequivocal on the role of Hox genes in
34 the specification of various body parts along the antero-posterior axis, the mechanisms by
35 which Hox proteins specifically identify and regulate the expression of their target genes
36 remain elusive. As expression, structure, and function of Hox proteins are comparable among
37 closely related species, it is postulated that variations in the regulation of targets of Hox
38 proteins are critical for morphological diversification.
39
40 Multiple monomeric and low affinity binding sites are reported to confer specificity to
41 Hox proteins (Galant et al. 2002; Crocker et al. 2015), which may also correlate to higher levels
42 of Hox protein needed to achieve robustness in their function (Paul et al, 2021). Cofactors of
43 Hox proteins such as TALE class of cofactors, particularly PBC and Meis family proteins, are
44 also shown to be involved in the regulation of Hox target gene expression (reviewed in
45 Merabet and Galliot 2015; Mann and Affolter 1998). However, various distal appendages in
46 arthropods and vertebrates do not require the input of either the PBC or Meis family of
47 cofactors which might be suggestive of a cofactor independent mechanism of target
48 regulation.
49
50 In *Drosophila melanogaster*, the Hox protein Ultrabithorax (Ubx) specifies the
51 development of halteres in the third thoracic (T3) segment (Lewis, 1978 and many other
52 reports; reviewed in Khan et al. 2020). Loss of function mutations of Ubx gives rise to four
53 winged flies with complete duplication of the T2 segment in place of T3 (Lewis 1978; Bender
54 et al. 2007), while overexpression of Ubx in larval wing discs lead to wing to haltere
55 transformations (Cabrera et al. 1985; White and Akam 1985; White and Wilcox 1985; Castelli-
56 Gair et al. 1990). Ubx is known to bind, in vitro, to motifs with consensus binding site
57 (TTTAATTA) having a TAAT core sequence (hereafter referred to as TAAT motif) (Ekker et al.
58 1991; Noyes et al. 2008; Berger et al. 2008). ChIP-Chip studies suggest that Ubx binds to a
59 large number of sites in the genome (Slattery et al. 2011; Choo et al. 2011; Agrawal et al.
60 2011). While most of those Ubx-binding sites have TAAT motif, they are not specifically
61 enriched in pulled down sequences compared to their prevalence in the whole genome
62 (Agrawal et al. 2011). Furthermore, many of those direct targets of Ubx are not differentially
63 expressed between developing wings and halteres (Agrawal et al. 2011) . One possibility is
64 due to its low-affinity binding to the chromatin, ChIP may pull down a very large number of

65 enhancers and not all of them are real targets of Ubx. Thus, it is not well understood what is
66 the mechanism by which Ubx specifically recognizes and regulates its targets in the context
67 of haltere specification.
68
69 Ubx is expressed in hindwing primordia of different insect species (Carroll 1995;
70 Prasad, Tarikere et al. 2016). For example, in the Coleopteran *Tribolium castaneum*, Ubx is
71 expressed and required for the specification of hindwings in the T3 (Tomoyasu et al. 2005),
72 while in Lepidopterans like *Precis coenia*, it specifies differences in eyespot patterns between
73 the hindwings and forewings (Weatherbee et al. 1999). In Hymenopterans like *Apis mellifera*
74 which bear a marginally smaller hindwing as compared to the forewing, Ubx is expressed in
75 both forewing and hindwing primordia. The expression, however, is stronger in the hindwing
76 primordia (Prasad, Tarikere et al. 2016). In *Bombyx mori*, although hindwing is
77 morphologically similar to the forewing, Ubx is expressed only in developing hindwings. Over-
78 expression of Ubx derived from *A. mellifera or B. mori* or *T. castaneum* can suppress wing
79 development and specify haltere fate in transgenic *D. melanogaster*, suggesting that changes
80 in Ubx at the protein levels may have contributed only marginally to the evolution of its
81 function. A study comparing genome-wide targets of Ubx in developing halteres of *D.*
82 *melanogaster* and developing hindwings of *A. mellifera* and *B. mori* reveal that a large number
83 of genes have remained common targets of Ubx between the two species. Among those
84 common targets, a few wing patterning genes are differentially expressed in *D. melanogaster*,
85 but not in *A. mellifera* or *B. mori* (Prasad, Tarikere et al. 2016). Evolutionary changes in cis-
86 regulatory elements are thought to be associated with this phenomenon.
87
88 In this study, we performed ChIP-sequencing for the Ubx protein in third instar haltere
89 imaginal discs from *D. melanogaster*. Our analysis led to the identification of a motif
90 (MATAAATCAY) containing a TAAAT core sequence (hereafter referred to as TAAAT motif)
91 that was enriched in ChIP pulled-down sequences. The prevalence of this motif, however, was
92 found to be similar between putative enhancers of direct targets of Ubx that are upregulated
93 or downregulated in the haltere. In vitro analyses suggest that Ubx binds strongly to TAAAT
94 motif than the previously reported TAAT motif. We experimentally analyzed enhancer of one
95 target that is upregulated (*CG13222*) by Ubx during haltere specification and one that is down

96 regulated (*vestigial*). Studies using the *CG13222* enhancer in *D. melanogaster* halteres
97 suggest that the TAAAT motif is functionally relevant and is critical for its upregulation by Ubx.
98
99 In *A. mellifera*, the enhancer of the *vestigial* (*vg*) gene has a TAAT motif, and no TAAAT
100 motifs. In transgenic *D. melanogaster,* this orthologous enhancer of *vg* of *A. mellifera* drives
101 reporter gene expression in both wing and halteres and is not sensitive to over-expression or
102 loss of Ubx. Changing the TAAT motif to TAAAT motif was sufficient to bring this, otherwise
103 unresponsive, enhancer under the regulation of Ubx. Comparative genome-level analyses
104 suggest that, unlike in *D. melanogaster*, TAAAT motifs are not enriched in ChIP-pulled down
105 sequences by Ubx from developing hindwings of *A. mellifera*. Taken together, our study
106 suggests that evolutionary changes in the regulatory sequences in the targets of Ubx in
107 dipteran lineages, specifically from TAAT to TAAAT as core sequence in Ubx-binding sites, is
108 correlated to the evolution of Ubx-mediated specification of haltere.
109
110 **Results**
111 **Motif with a TAAAT core sequence is enriched in direct targets of Ubx in developing halteres**
112 **in *D. melanogaster***
113 Several previous studies, many from our own group, have reported identification of direct
114 targets of Ubx in developing halteres. Those studies employed single gene approaches or
115 ChIP-chip based methods. In order to identify Ubx-binding sites with higher accuracy and
116 resolution in the genome, we employed ChIP-sequencing, using antibodies that are specific
117 to Ubx to pull down the chromatin from the third instar haltere discs. Our analysis revealed
118 2171 high confidence peaks for Ubx annotating to 1236 direct targets, of which 418 genes
119 were common to ChIP-microarray data from Choo. et al, 2011 and 148 genes were common
120 to Agrawal et al, 2011 (Fig. S1 A). We observed greater occupancy of Ubx at intronic and
121 intergenic regions (Fig. S1 B). Several previously reported direct targets of Ubx such as
122 *CG13222* and *spalt* were also part of the ChIP-seq data (Fig. S1 C, S1 D).
123
124 We carried out extensive bioinformatic analysis to identify common sequence motif/s
125 in pulled down sequences, which is/are expected to be recognition sites for Ubx to bind to its
126 targets. Using a de-novo motif finding algorithm, we observed that a motif (MATAAATCAY)
127 with TAAAT core sequence is enriched in Ubx pulled down sequences. This motif was found

128 to be 1.7fold overrepresented in Ubx pulled down sequences as compared to its prevalence
129 in the whole genome (Fig. 1B). The canonical TAAT motif is not enriched in Ubx-pulled down
130 sequences, rather it was found under-represented compared to its prevalence in the whole
131 genome (Fig. 1B). Similar observations are reported earlier when ChIP-chip method was
132 employed to identify direct targets of Ubx (Agrawal et al., 2011).
133
134 Earlier studies have reported that TAAAT motif is bound by the Ubx-Extradenticle (Exd)
135 complex during embryonic development. However, Exd is not required for the development
136 of the haltere capitellum (Casares and Mann 2000). In this context, we first sought to
137 determine functional significance of Ubx-binding to the TAAAT motif. We carried out RNA-
138 seq analyses to identify genes that are differentially regulated between the third instar wing
139 and haltere imaginal discs (generation of RNA-seq and analyses- unpublished data). Among
140 611 genes that are differentially expressed between wing and haltere discs, 71 genes are
141 common to ChIP-Seq data suggesting them to be potential direct targets of Ubx. We
142 categorised direct targets of Ubx expressed in wing and/or haltere discs into three categories:
143 genes that are upregulated in haltere discs (44), downregulated in haltere discs (27) and genes
144 that are expressed at equal levels between wing and haltere discs (99) (Fig. 1C). We found
145 that the frequency of the TAAAT motif in putative Ubx response elements in these three
146 categories of targets are similar (Fig. 1D). While TAAAT motif may be preferred site for Ubx to
147 bind on the enhancers of its targets, it was not obvious from these analyses its significance in
148 the context of Ubx-mediated regulation of its targets.
149
150 **The TAAAT motif in the CG13222 enhancer confers high-affinity and specificity in Ubx**
151 **binding *in-vitro***
152 We next evaluated relative affinities and specificities of Ubx binding to the TAAAT motif to
153 regulate its targets as against binding to the TAAT motif. We chose two well-characterized
154 targets of Ubx, Spalt (*sal*) and *CG13222*. While Ubx downregulates *sal* (Galant et al. 2002), it
155 upregulates *CG13222* (Mohit et al. 2006; Hersh et al. 2007). The enhancer of *sal* contains
156 several sites with the TAAT motif (Fig. 2A) and these sites were shown to be recognized by
157 Ubx for effective repression in the haltere imaginal disc (Galant et al. 2002). The "*edge*"
158 enhancer of *CG13222* has two TAAT motifs at 17 bp apart. Hersh et al. (2007) have named the
159 two motifs as site 1 and site 2. In their transgenic assays, site 1 appears to be critical for Ubx-

160 mediated regulation of *CG13222*, while site 2 is dispensable. Interestingly, the site1 in the
161 *edge* enhancer which is critical for its regulation by Ubx has a TAAAT motif in proximity to the
162 TAAT motif (Fig. 2A). Their analysis of site1 involved mutating both TAAT and TAAAT motifs,
163 which leads to loss of enhancer driven reporter gene expression (Hersh et al. 2007). We
164 mutated site1 in such a way that it affected only the TAAAT (M1_A) or the TAAT motif (M1_B)
165 (Fig. 2A) and examined the relative affinities of binding of Ubx to these sites.
166
167 First we directly estimated binding affinities by employing Electrophoretic mobility
168 shift assays (EMSA). We used a 29bp long probe for *sal* and a 21bp long probe covering site 1
169 of *edge* enhancer of *CG13222* for EMSA. Two different Hox proteins, Ubx and Sex combs
170 reduced (Scr) were tested on these probes for assessing DNA-binding specificity *in vitro*.
171 Results showed that both the Hox proteins bind on the probe derived from the *sal* enhancer,
172 although with less affinity for Scr (probe Spalt 5/6 in Fig. 2B). In contrast, only Ubx could bind
173 on the probe derived from the *CG13222 edge* enhancer suggesting a higher level of DNA-
174 binding selectivity when compared to the *sal* enhancer (probe WT (site1) in Fig. 2B). We find
175 that the mutation in TAAAT motif in site1 of *CG13222* led to a complete loss of Ubx binding,
176 while the mutation in TAAT motif had no significant effects (probes M1_A and M1_B,
177 respectively in Fig. 2B). These experiments, thus, demonstrate that the TAAAT motif may
178 confer higher affinity and specificity for Ubx to bind on the enhancer of *CG13222* (Fig. 2B).
179
180 Next we examined the relative importance of the two motifs in gene regulation using
181 Luciferase reporter assays in S2 cells. For these functional assays, we used a 512bp long region
182 spanning the *edge* enhancer of *CG13222* as reported by Hersh et al. (2007), which has both
183 site 1 (consisting of one TAAT motif and one TAAAT motif) and site 2 consisting of one TAAT
184 motif). Consistent with *CG13222* being a target of Ubx upregulated during haltere
185 development, a luciferase expression construct driven by the enhancer of the gene was
186 upregulated in S2 cells only in the presence of Ubx (construct WT in Fig. 2C and Fig. S2 A). We
187 observed a significant decrease in Ubx-dependent expression compared to the wildtype
188 *CG13222* enhancer when TAAAT alone is mutated (construct M1_A in Fig. 2C), thus suggesting
189 that TAAAT motif is essential for Ubx-dependent expression of *CG13222*. We also attempted
190 to examine the effect of mutating the TAAT motif to a TAAAT motif (in site2), which results in
191 two high-affinity binding motifs in the *edge* enhancer of *CG13222*. We observed marginal

192 reduction in the luciferase activity (Fig. S2 B), perhaps, due to negative effect of two high-
193 affinity binding sites for Ubx in close proximity (however see below results of transgenic
194 assays).
195
196 **TAAAT motif is essential for Ubx-mediated activation of CG13222 *in vivo***
197 To assess the role of the TAAAT motif in Ubx-mediated activation of *CG13222* enhancer in the
198 haltere imaginal discs, we generated transgenic strains of *D. melanogaster* carrying the wild
199 type or mutant enhancers cloned upstream of a GFP reporter. The wildtype *CG13222*
200 enhancer showed GFP expression at the posterior edge of the haltere imaginal discs
201 consistent with earlier reports (Fig. 2D; Mohit et al. 2006; Hersh et al. 2007). We generated
202 transgenic lines for multiple mutant versions of the *CG13222* enhancer. The reporter gene
203 expression is significantly reduced when driven by the M1 mutant enhancer, which mutates
204 both TAAT and TAAAT motifs at site 1 of the enhancer (Fig. 2E). We observed similar loss of
205 reporter expression when the GFP was driven with a version of mutant enhancer (M1_A
206 mutant) which mutates only the TAAAT motif (Fig. 2F, 2H). This suggests that Ubx mediated
207 activation of the *CG13222* enhancer is dependent on the TAAAT motif.
208
209 As described earlier, the *edge* enhancer of *CG13222* has a second TAAT motif (site 2),
210 which is redundant for Ubx-mediated upregulation of *CG13222* in the haltere discs (Hersh et
211 al. 2007). We generated a transgenic line (M2_A) wherein this TAAT core sequence is mutated
212 to TAAAT core sequence resulting in two TAAAT motifs in the modified enhancer as against
213 one in the wildtype. This transgene showed much stronger expression of the GFP reporter
214 compared to the wildtype *CG13222* enhancer. Enhancer driven expression was also observed
215 in ectopic regions in the posterior compartment (red arrows in Fig. 2G), further suggesting
216 that the presence of a TAAAT motif brings the, otherwise redundant, site2 under Ubx
217 regulation.
218
219 **Introduction of a TAAAT motif is sufficient to bring an otherwise unresponsive enhancer of**
220 ***vestigial* gene of *A. mellifera* under Ubx control in a transgenic *Drosophila* assay**
221 Examination of *CG13222* enhancer suggested an essential role for TAAAT motifs in Ubx
222 mediated upregulation of target genes. We next analysed cis-regulatory regions of *vg*
223 (*vestigial*), a pro-wing gene, which is downregulated by Ubx in developing halteres,

224 (Weatherbee et al., 1998; Shashidhara et al., 1999) for the presence of TAAAT sites. Of the
225 two well-characterized enhancers of *vg*, the quadrant enhancer (hereafter referred to as
226 *quad-vg* enhancer) is differentially expressed between wing and haltere discs. Scanning for
227 sequence motifs revealed that this enhancer has a 25bp cassette containing both TAAAT and
228 TAAT motifs in close proximity (Fig. 3A). Sequence alignment of the region containing the Ubx
229 binding motifs show high conservation of both TAAT and TAAAT among most *Drosophila*
230 species (Fig. S3 A).
231
232 We examined the relative importance of TAAAT and TAAT motifs for Ubx-dependent
233 downregulation of *quad-vg* enhancer in a transgenic assay. As reported earlier, we observed
234 that the wild type *quad-vg* enhancer drives expression of the GFP reporter in the wing
235 imaginal discs, albeit at low levels (compared to earlier reported *quad-vg* -lacZ reporter), but
236 not in the haltere imaginal discs (Fig 3B). To test a possible role for the TAAAT motif in
237 mediating differential expression of this enhancer, we generated transgenic flies carrying
238 mutations (Fig. S3 B) in the Ubx-binding sites. However, we observed complete loss of
239 reporter GFP expression in both wing and haltere imaginal discs for most mutant constructs
240 of *quad-vg*, suggesting that mutations in Ubx-binding motifs of the *quad-vg* enhancer may
241 completely inactivate the enhancer element (Fig. 3B, Fig. S3 C).
242
243 We had previously reported that the *vg* gene in *A. mellifera* (hereafter referred to as
244 *Apis-vg*) is expressed in both fore- and hindwings (Prasad, Tarikere et al. 2016). In a transgenic
245 *Drosophila* assay, an enhancer of *Apis-vg* equivalent to *quad-vg* showed similar patterns and
246 levels of expression between wing and haltere discs (Prasad, Tarikere et al. 2016). However,
247 we could not locate a specific region in the two enhancers that is relevant to understand the
248 evolution of differential expression of *vg* in wing and haltere in *D. melanogaster*. We
249 specifically searched for TAAAT motif/s in the enhancer of *Apis-vg*. While enhancer of *Apis-*
250 *vg* contains a TAAT motif, we did not find any TAAAT motifs (Fig. 3C). We tested the
251 significance of this in a transgenic *Drosophila* assay. We generated new transgenic flies with
252 wild type and mutant enhancers of *Apis-vg*. As reported earlier, we observed similar and
253 strong expression of GFP in both wing and haltere imaginal discs (Fig. 3D). *vg* gene in *D.*
254 *melanogaster* is expressed in both the presumptive hinge and pouch regions. However, the
255 *quad-vg* enhancer of *D. melanogaster* drives its expression only in the non dorso-ventral

256 boundary cells of the pouch. *Apis-vg* driven GFP expression, while was much lower in the
257 dorso-ventral boundary cells, is at equivalent levels in both the presumptive hinge and the
258 pouch regions. The expression pattern of GFP driven by the enhancer of *Apis-vg* in *Drosophila*
259 wing discs is an indication of high degree of conservation of this enhancer between *A.*
260 *mellifera* and *D. melanogaster* (which have diverged for more than 350 million years). Wing
261 patterning genes upstream of *vg* are able to regulate *Apis-vg* enhancer in a pattern similar to
262 its counterpart in *D. melanogaster*. Nonetheless, this enhancer is not sensitive to Ubx-
263 mediated regulation of gene expression in haltere discs.
264
265 We generated a number of transgenic flies carrying various mutations in the *Apis-vg*
266 enhancer with the aim to understand the significance of the absence of TAAAT motif vis-à-vis
267 its expression in haltere discs in transgenic *Drosophila*. First we generated a chimeric
268 enhancer (*Apis-vg_M2*), in which a 25bp Ubx-binding cassette from the *Drosophila quad-vg*
269 enhancer replaced its counterpart in the *Apis-vg* enhancer (Fig. S4 A (iii)). *Apis-vg_M2* showed
270 differential expression of GFP reported between wing and haltere imaginal discs (Fig. S4 C).
271 Its expression was much lower in the haltere pouch compared to the GFP reporter driven by
272 the wildtype *Apis-vg* enhancer (Fig. S4 C). In another chimeric enhancer, only a part of the
273 cassette with the TAAAT motif was introduced from *quad-vg* of *D. melanogaster* (*Apis-*
274 *vg_M3*). This too resulted in differential expression of the GFP reporter in wing and haltere
275 discs (Fig. S4 C). Interestingly, the GFP expression was lower only in the pouch region and not
276 in the hinge region suggesting that these manipulations of the sequence of *Apis-vg* enhancer
277 brings its expression patterns closer to that of *quad-vg* of *Drosophila*. More importantly,
278 changing only the TAAT motif of the *Apis-vg* enhancer to a TAAAT motif (*Apis-vg_M1*) was
279 sufficient for the repression of the GFP reporter in the haltere pouch without affecting its
280 expression in the wing pouch (Fig. 3D, 3E, 3F, Fig. S4 B). This suggests that micro-evolutionary
281 changes in the enhancer sequence may have brought the *vg* gene under the regulation of Ubx
282 in dipteran lineages. This gene being a pro-wing gene, its repression is critical for Ubx-
283 mediated specification of haltere.
284
285 **Enrichment of TAAAT motif in the enhancers of targets of Ubx is specific to *D. melanogaster*:**
286 Our results suggest that differences in Ubx binding motif (TAAAT vs TAAT) in enhancers of
287 orthologous targets may have brought key wing patterning gene *vg* under the differential

288 regulation by Ubx in dipteran lineage. To understand whether the difference in Ubx binding
289 motif might be a recurrent theme that impinges on differential hindwing morphology, we
290 compared genome wide binding of the Ubx protein between *D. melanogaster* and *A.*
291 *mellifera*.
292
293 Previous studies indicate that despite the differences in the hindwing morphology, the
294 Ubx protein is expressed in the third thoracic segment of both insects (as well as other insect
295 lineages) (Fig. 4A) (Carroll 1995; Prasad, Tarikere et al. 2016). Additionally, Ubx is highly
296 conserved across *A. mellifera, B. mori, T. castaneum, Junonia coenia* and *D. melanogaster* at
297 the level of its DNA binding homeodomain and the protein interaction domains, YPWM and
298 Ubda (Fig. S5 A). We have previously shown that overexpression of Ubx from different insect
299 species in the wing imaginal disc of transgenic *Drosophila* resulted in wing to haltere
300 transformation in a pattern similar to *Drosophila* Ubx (Prasad, Tarikere et al. 2016). This
301 suggests that differences in insect hindwing morphology is a consequence of events occurring
302 downstream or parallel of Ubx. Comparison of targets of Ubx between *D. melanogaster* and
303 *A. mellifera* suggest that many functionally important genes are common to the two lineages
304 (Fig. S5 B). This reconfirms that morphological differences may lie at the levels of regulation
305 of those targets.
306
307 We analysed previously reported Ubx ChIP-seq data of *A. mellifera* (Prasad, Tarikere
308 et al. 2016) for the presence of TAAT and TAAAT motifs. We did not observe any enrichment
309 of either TAAT or TAAAT motifs in the *A. mellifera* Ubx ChIP-seq dataset (Fig. 4B, 4C). However,
310 we observed enrichment for binding sites for other transcription factors GAGA, CTCF and Brk
311 in both *D. melanogaster* and *A. mellifera* datasets (Fig. 4B). Interestingly, the frequency of the
312 TAAAT motif in the entire genome is 1.8 fold higher in *A. mellifera* as compared to *D.*
313 *melanogaster* (Fig. 4C). This suggests that the TAAAT motif is selectively enriched in the
314 enhancers of targets of Ubx only in *D. melanogaster*.
315
316 In addition, we also observed differences in the clustering of peaks vis-à-vis
317 transcription start site (TSS) in the two species (Fig. S5 C). While in *D. melanogaster*, Ubx
318 binding sites are clustered away from the TSS, the reverse is observed in *A. mellifera*.

319 However, further studies are needed to understand functional relevance, if any, of this
320 topology of binding sites.
321
322 Taken together, In the context of the fact that Ubx brings differential development of
323 the dorsal appendages of T2 and T3 in *D. melanogaster*, but not in *A. mellifera*, our results
324 help uncover a mechanism in which micro-evolutionary changes in enhancer sequences could
325 bring in key wing patterning genes under the regulation of Ubx.
326
327 **Discussion**
328 Genetics and developmental biology methods have established, unequivocally, the role of
329 Hox genes as master control genes in regulating segment-specific developmental pathways.
330 However, precise mechanism by which each of the Hox proteins mediate development of a
331 specific morphological feature is still largely debated. Main reasons this has eluded systematic
332 study include their highly conserved DNA-binding domains in both Hox and non-Hox
333 homeobox-containing transcription factors and their overlapping expression patterns in a
334 given set of cells.
335
336 In *D. melanogaster*, Ubx activates and represses a number of genes at various levels
337 of hierarchical wing patterning pathways, some by direct regulation and largely by indirect
338 regulation. Identifying mechanisms by which Ubx directly regulates a few key wing patterning
339 genes may throw a light on the mechanism by which it specifies haltere development. For
340 example, in the context of unequal size of the anterior and posterior compartments in haltere
341 discs (as against equal size in wing discs), multiple studies have elucidated an elegant
342 mechanism (Mohit et al. 2006; de Navas et al. 2006; Crickmore and Mann 2006; Makhijani et
343 al. 2007; Crickmore and Mann 2007). Ubx activates the expression of Thickvein (Tkv) at the
344 A/P boundary and suppresses the expression of Dally in the posterior compartment making
345 the gradient of growth molecule Decapentaplegic (Dpp) more pronounced in the anterior
346 compartment than the posterior compartment.
347
348 Efforts over the past 3 decades from multiple laboratories, including ours, are centred
349 around the hypotheses that Ubx-mediated specification of haltere development involves (i) a
350 possible unique Ubx-binding motif that identifies a gene as a target of Ubx; (ii) modulation of

351 expression of a set of direct targets would lead to distinct changes in number, size, shape, and
352 organization of epithelial cells resulting haltere structure as against wing structure and (iii)
353 step-wise regulation of wing developmental pathways to modify its fate to haltere.
354
355 Earlier efforts involving whole genome approaches, such as microarray and ChIP-chip
356 analyses, to identify all potential targets of Ubx have not yielded satisfactory results. With the
357 advent of higher resolution, specificity and sensitivity in genome sequencing techniques, we
358 envisaged ChIP-Seq and RNA-seq to identify direct targets of Ubx that are functionally
359 relevant to haltere specification. Detailed lists of all genes identified using these methods and
360 the implications of the same to understand the mechanism of function of Ubx would be
361 published elsewhere. While analysing the enhancer regions of direct targets of Ubx identified
362 by ChIP-Seq, we observed presence of a putative Ubx-binding motif with TAAAT core
363 sequence, which is different from the one (TAAT motif) being studied for more than 2 decades
364 now. The fact that Ubx binding to this motif is stronger suggests a possible mechanism for
365 target identification. While low-affinity binding brings Ubx on the enhancers of a very large
366 number of genes, only those enhancers which have stronger affinity for Ubx would be
367 impacted at the functional level. Earlier reports suggest that specificity in gene regulation by
368 Hox proteins may come from stronger binding to their targets, which in turn could be ascribed
369 to higher levels of a given Hox protein in a specific segment (Paul et al. 2021) and/or to the
370 cofactors binding to nearby locations on the chromatin (Agrawal et al. 2011). Our observation
371 on the presence of high-affinity binding sites on the enhancers of targets of Ubx adds another
372 dimension to the mechanisms by which robustness is achieved during Ubx-mediated haltere
373 development.
374
375 We studied the consequence of binding of Ubx to TAAAT motif in the context of
376 regulation of two different genes. *CG13222*, a gene upregulated by Ubx in the haltere and *vg*,
377 a gene downregulated by Ubx. Our observations reported here suggest that Ubx binds
378 strongly to TAAAT motif than to TAAT motif and its binding to former is critical for the
379 regulation of *CG13222* expression. We also report that TAAT motif present in the enhancer
380 of *CG13222*, which is also ubiquitously present in the genome, is dispensable for Ubx-
381 mediated regulation.
382

383 However, we could not directly study Ubx-mediated regulation of *vg* in *Drosophila*,
384 perhaps due to Ubx binding to a region which is critical for its expression in developing wings.
385 As Ubx represses wing development to promote haltere development, this suggests a
386 possible mechanism in which Ubx interferes with gene regulatory network to activate *vg*
387 during wing development. Binding of Ubx with high affinity to this part of the enhancer of *vg*
388 may interfere with the binding of other transcription factors. This, however, needs to be
389 empirically validated.
390
391 An earlier report of ours (Prasad, Tarikere et al. 2016) and this report suggest that an
392 enhancer of *vg* gene of *A. mellifera* can drive GFP expression in wing imaginal discs in a
393 transgenic assay in a pattern similar to that of the quadrant enhancer of *Drosophila vg*. The
394 enhancer of *Apis-vg*, however, is not sensitive to Ubx and drives reporter gene expression at
395 high levels in the haltere. The enhancer of *Apis-vg* allowed us to dissect the importance of
396 TAAAT motif as we could achieve its repression in haltere discs without affecting its
397 expression in wing discs. We show that changing the TAAT motif to TAAAT motif, dramatically,
398 brought the enhancer under the negative regulation of Ubx. As *vg* is a pro-wing selector gene
399 and can assign wing fate to any group of dorsal epithelial cells (Kim et al. 1996; Klein and Arias
400 1998; Williams et al. 1991; Neumann and Cohen 1996), it is a critical target of Ubx to specify
401 haltere development. In this context, a microevolutionary change (TAAT to TAAAT motif) in
402 the enhancer of an ancestral *vg* may be a critical step in the evolution of dipterans.
403
404 We have observed that, in general, TAAAT motif is more prevalent among the direct
405 targets of Ubx in *D. melanogaster* than in *A. mellifera*. The *Drosophila* Ubx protein has longer
406 linker region (than the Ubx of *A. mellifera*) between the YPWM motif (which mediates
407 protein-protein interactions) and the homeodomain (which mediates binding to DNA) in the
408 isoform Ia that mediates haltere specification. Once Ubx binds strongly to the chromatin with
409 the help of TAAAT motif, it may interfere with other proteins in the neighbourhood modifying
410 the gene expression pattern from wing type to that of haltere. Thus, evolutionary changes in
411 the protein structure as well in sequences in the enhancers of its target genes may both be
412 involved in the evolution of function of Ubx as a modifier of wing development, re-specifying
413 its fate to haltere. It is, however, very unlikely that a large number of convergent mutations
414 would be involved in this process, wherein multiple TAAT motifs are evolved to TAAAT motifs.

415 A few critical genes, such as *vg*, upstream of wing patterning pathways may have evolved to
416 become targets of Ubx through this route. Once regulatory networks of those genes are
417 modulated by Ubx, chromatin landscape of many other targets (Ubx may bind to these targets
418 via TAAT motifs) would change making them amenable for Ubx-mediated regulation and
419 thereby specifying the haltere fate.
420
421 **Materials and Methods**
422
423 **1. ChIP sequencing:** The protocol was essentially as reported by (Agrawal and Shashidhara
424 2014). Third instar wandering larvae were cut, inverted and fixed with 1.5% PFA
425 (ThermoFisher scientific) for 20 minutes at room temperature and subsequently quenched
426 (125mM Glycine solution, 1XPBS). Samples were washed twice with 1XPBS (Sigma) for
427 10minutes each and dissected for wing and haltere discs (500 number of wing discs and 1000
428 number of haltere discs were used per replicate). Samples were lysed with Cell Lysis Buffer
429 (10mM Tris-Cl pH 8, 10mM NaCl, 0.5% NP40, 1xPIC (Roche)) followed by mechanical shearing
430 for 10minutes on ice. After centrifugation at 2000rpm for 10mins at 4 ^{o}C, the was pellet
431 resuspended in Sonication buffer (50mM Tris-Cl pH 8, 1% SDS, 1% NP40, 10mM EDTA, 1xPIC)
432 followed by incubation on ice for 30 minutes. Sonication was carried out using the Covaris S2
433 sonicator using settings (DC 20%, Intensity 5, Cycles of Burst 200, Time= 20 minutes). Samples
434 were centrifuged at 14000rpm for 15mins at 4 ^{o}C and precleared using with 4ml of Magnetic
435 A Beads (Diagenode) at 4 ^{o}C for 1 hour. Preclearing beads were removed and 2ml of Polyclonal
436 anti-Ubx antibodies (Agrawal et al. 2011) were added with overnight incubation at 4 ^{o}C.
437 Magnetic A beads were added and incubated for 4 hours at 4 ^{o}C. The supernatant was
438 removed and beads washed twice with low salt buffer (20mM Tris-Cl pH 8, 150mM NaCl, 0.1
439 % SDS, 2mM EDTA, 1% TritonX-100), twice with High Salt buffer (20mM Tris-Cl pH 8, 500mM
440 NaCl, 0.1 % SDS, 2mM EDTA, 1% TritonX-100), twice with LiCl buffer (10mM Tris-Cl pH 8,
441 250mM LiCl, 1% NP40, 1% Na-Deoxycholate, 500mM EDTA,), and once with TE buffer (10mM
442 Tris-Cl pH 8, 500mM EDTA). Samples were eluted in elution buffer (50mM Tris-Cl, 1mM EDTA,
443 1% SDS, 50mM NaHCO3). Samples were decrosslinked followed by treatment with RNAse
444 treatment at 37 degrees for 1hour and Proteinase K treatment at 42 ^{o}C for 2 hours. DNA was
445 purified using PCI purification and quantified using the Qubit HS DNA quantification system.
446

447 For library preparation, equal amount of DNA (~2 ng) was used as an input for NEB
448 Ultra II DNA library prep kits (NEB #E7645). Number of cycles for amplification of adapter
449 ligated libraries were estimated by the qPCR before final amplification to avoid any bias
450 arising due to PCR amplification and indexing (NEB #E7350). Final amplified libraries were
451 purified twice, first with 1X followed by 0.8x volume of beads per sample using HiPrep PCR
452 clean up system (Magbio #AC-60050). Library concentration was determined using Qubit HS
453 DNA kit (Invitrogen #E7350) and average fragment size was estimated using DNA HS assay on
454 bioanalyzer 2100 (Agilent #5067-4626) before pooling libraries in equimolar ratio. Sequencing
455 reads (100bp PE) were obtained on the Hiseq 2500 V4 platform at Macrogen Inc, Korea.
456
457 **2. ChIP seq analysis:** For *D. melanogaster*, raw reads were trimmed using Trimmomatic
458 software and aligned to the dm6 genome (version BDGP6.28). Peak calling was performed
459 using MACS2 (maxdup=1, FDR 0.05) (Zhang et al. 2008) and high confidence peaks (occurring
460 in at least two biological replicates) were identified (https://ro-che.info/articles/2018-07-11-
461 chip-seq-consensus).
462
463 For *Apis mellifera*, raw reads were downloaded from NCBI (GSE71847). Reads were
464 aligned to the Amel_4.5 genome (version 4.5.47) and peak calling done using MACS2
465 (maxdup=1, FDR 0.05). High confidence peaks occurring in both replicates were identified.
466
467 Peaks were annotated to their nearest TSS using the Homer software
468 (annotatePeaks.pl) (Heinz et al. 2010). Gene ontology of targets was performed using GO
469 (Ashburner et al. 2000; Carbon et al. 2021) and overrepresented GO terms plotted. The
470 ChIPseeker program (Yu et al. 2015) was used to annotate peaks to genomic regions. Genome
471 tracks were made using Deeptools (normalizeUsing RPGC, binsize 1, minMappingQuality 10,
472 smoothLength 150) (Ramírez et al. 2016) and visualized using the IGV software (Robinson et
473 al. 2011). Heatmaps were generated using Deeptools (referencePoint TSS).
474
475 **3. Motif analyses and protein sequence comparison:** De-novo motif analysis was carried out
476 using Homer (findMotifsGenome.pl). The PWM for the TAAAT motif obtained from Homer
477 was converted to Transfac format using RSAT (Nguyen et al. 2018) and finally converted to
478 MEME format using transfac2meme command. A 100bp pad on both sides was applied to

479 ChIP peaks for calculation of frequency of motifs usingthe FIMO software from MEME suite.
480 Identification of binding motifs in enhancers of *CG13222, quad-Vg* and *ApisVg* was performed
481 using the MAST software from MEME suite (Bailey et al. 2009).
482
483 Protein sequences of Ubx from different insect species were downloaded from
484 Uniprot and multiple sequence alignment performed using Clustal Omega (Madeira et al.
485 2019). Alignments were visualised and processed using the Jalview software (Waterhouse et
486 al. 2009).
487
488 **4. Comparative analyses of putative Ubx response elements:** Genes that are differentially
489 expressed between the wing and haltere were identified using RNA sequencing of third instar
490 wing and haltere imaginal discs (unpublished data). A fold difference (FD) cut-off of 1.5 fold
491 was applied to identify upregulated or downregulated genes in the haltere. Genes having a
492 fold difference ranging between 0.95-1.05 between the wing and haltere imaginal discs were
493 considered as not-differentially expressed. Ubx targets obtained from ChIP-seq was
494 compared to RNA-seq data and frequency of the TAAAT motif calculated using an analysis
495 module (unpublished data).
496
497 **5. Plasmids constructs, cloning and transgenic fly generation:** The full list of primers as well
498 as sequences of enhancers of the *CG13222, quad-Vg* and *ApisVg* is provided as supplementary
499 data (File S2). For Luciferase assays, site-directed mutagenesis was used to clone enhancer
500 constructs of the *CG13222* gene between Kpn1 and Nhe1 restriction sites, upstream of a
501 modified pGL3 vector containing a 5X Dorsal binding site. Metallothionein inducible pRMHa3
502 vectors containing Ubx were previously generated in the lab and empty pRMHa3 vector was
503 generated by excising the cloned Ubx sequence.
504
505 For generating transgenics, all enhancer constructs were cloned between the Nhe1
506 and Kpn1 restriction sites of the pH-stinger-attb (a kind gift from Manfred Frasch) and injected
507 into the attp40 site on the second chromosome using a mini white screen at NCBS fly facility,
508 Bangalore. All constructs were sequence verified before use in functional assays.
509

510 **6. Luciferase assays:** S2 cells were checked for contamination before performing Luciferase
511 assays. Cells were plated onto 24 well plates at a density of 3*10^5 cells per well 6 hours prior
512 to transfection. For every construct, either wild type or mutant, two sets of experiments were
513 designed; one well was co-transfected with the enhancer construct in pGL3 vector and the
514 empty pRMHa3 vector whereas the other well was co-transfected with the enhancer
515 construct and pRMHa3 vector containing Ubx. Renilla luciferase was used as an internal
516 control and co-transfected in all experiments. Transfection was carried out using the
517 Effectene transfection reagent and all experiments carried out in 3 technical replicates and at
518 least 3 biological replicates. 48 hours post transfection, sterile CuSO4 solution was used to
519 induce expression of Ubx at a final concentration of 500um and incubated for 24 hours. Cells
520 were harvested, pelleted down (1000rpm for 4mins) and 100ul of 1X Passive Lysis buffer
521 added and vortexed to dissolve the cell pellet. Cells were incubated for 15mins at room
522 temperature and further spun at 10,000rpm for 90secs to collect the supernatant. The
523 luminescence was measured with the Dual-glo Luciferase assay kit (Promega) on Ensight Plate
524 reader (Perkin Elmer). All readings were normalized to Renilla luminescence and datasets
525 compared using the Prism software.
526
527 **7. EMSA:** VC-Ubx and VC-Scr were cloned in the PcDNA3 vector and produced with the TNT-
528 T7-coupled in vitro transcription/translation system (Promega) for EMSAs, as previously
529 described (Hudry et al. 2012). Shortly, between 3 and 6 microliters of programmed lysate was
530 used for each protein (100 ng/ml of proteins were produced on average). The VC-Ubx and VC-
531 Scr fusion proteins were produced separately (0,5mg of each plasmid was used for the in
532 vitro transcription/translation reaction). Supershift against the VC fragment was performed
533 by adding the anti-GFP antibody after 15 minutes in the binding reaction. Each band shift
534 experiment was repeated at least two times.
535
536 **8. Immunostaining and Microscopy:** Wandering third instar larvae were cut and inverted in
537 cold PBS followed by fixation with 4% PFA for 20mins with gentle rocking. Samples were
538 washed thrice with 0.1% PBTX (0.1% Triton in PBS) for 10 minutes each, followed by one hour
539 of blocking at room temperature (blocking solution: 0.5% BSA in 0.1%PBTX). Samples were
540 incubated with primary antibodies (dilutions made in blocking solution) at 4 ^{o}C overnight
541 followed by washing with 0.1% PBTX for 10mins (x3) at room temperature. Samples were

542 incubated with secondary antibodies for one hour at room temperature, washed with 0.1%
543 PBTX for 10mins (x3) and finally with PBS before mounting. Wing and haltere imaginal discs
544 were dissected out, mounted using Prolong Gold Antifade (Invitrogen) and stored at 4
545 degrees. Antibodies used in the study are Rb-GFP (1:1000) (Invitrogen), Rb-Ubx (1:1000)
546 (Agrawal et al., 2011), m-Ubx (1:30) (DSHB), goat-Hth (1:100) (Santa Cruz Biotechnology),
547 Alexa Fluor 568 (1:1000) (Invitrogen) and Alexa Fluor 633 (1:1000) (Invitrogen). All images
548 were taken on Leica Sp8 system at Microscopy Facility, IISER Pune. Laser settings (Laser Power
549 (LP) and Gain (G)) used for measuring GFP fluorescence have been indicated for each image
550 panel. All images were processed using the ImageJ software and compiled using Microsoft
551 PowerPoint and Adobe Illustrator.
552
553 **Data availability**: Raw fastq files of ChIP sequencing for the Ubx protein in *Apis mellifera*
554 hindwings can be accessed from NCBI (GSE71847). Raw data for ChIP sequencing for Ubx in
555 *Drosophila melanogaster* has been submitted to GEO. RNA sequencing data for the wing and
556 haltere imaginal discs will be published elsewhere.
557
558 **Acknowledgments**
559 We thank G. Deshpande and members of the LSS and SM laboratories for critical input. We
560 thank S. Galande for providing reagents for Library preparation. This work was supported
561 primarily by an Indo-French Research grant from CEFIPRA to SM; a JC Bose Fellowship and
562 grant from the Department of Science and Technology, Government of India to LSS; and a
563 Council of Scientific & Industrial Research (CSIR) Fellowship to SK.
564
565 **Author contributions:**
566 SK carried out all fly experiments, image analyses, and wrote the MS.
567 SK and SP did ChIP-Seq and RNA-Seq and analyses.
568 GG, FB and RP did EMSA experiments.
569 LSS and SM conceived the project and wrote the MS.
570 We declare “no-conflict-of-interest”.

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.

Agrawal P, Shashidhara LS. 2014. ChIP for Hox proteins from Drosophila imaginal discs. *Methods Mol Biol* **1196**: 241–253.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. 2000. Gene Ontology: tool for the unification of biology. *Nat Genet* **25**: 25–29. https://doi.org/10.1038/75556.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res* **37**: 202–208.

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.

Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Peña-Castillo L, Alleyne TM, Mnaimneh S, Botvinnik OB, Chan ET, et al. 2008. Variation in Homeodomain DNA Binding Revealed by High-Resolution Analysis of Sequence Preferences. *Cell* **133**: 1266–1276.

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.

Carbon S, Douglass E, Good BM, Unni DR, Harris NL, Mungall CJ, Basu S, Chisholm RL, Dodson RJ, Hartline E, et al. 2021. The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Res* **49**: D325–D334.

Carroll SB. 1995. Homeotic genes and the evolution of arthropods and chordates. *Nature* **376**: 479–485.

Casares F, Mann RS. 2000. A dual role for homothorax in inhibiting wing blade development and specifying proximal wing identities in Drosophila. *Development* **127**: 1499–1508.

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.

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.

Crickmore MA, Mann RS. 2007. Hox control of morphogen mobility and organ development

through regulation of glypican expression. *Development* **134**: 327–334.

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

de Navas LF, Garaulet DL, Sánchez-Herrero E. 2006. The Ultrabithorax Hox gene of Drosophila controls haltere size by regulating the Dpp pathway. *Development* **133**: 4495–4506.

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.

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.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* **38**: 576–589. http://dx.doi.org/10.1016/j.molcel.2010.05.004.

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.

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.

Hudry B, Remacle S, Delfini MC, Rezsohazy R, Graba Y, Merabet S. 2012. Hox proteins display a common and ancestral ability to diversify their interaction mode with the pbc class cofactors. *PLoS Biol* **10**.

Khan S, Dilsha C, Shashidhara LS. 2020. Haltere development in D. melanogaster: Implications for the evolution of appendage size, shape and function. *Int J Dev Biol* **64**: 169–175.

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.

Klein T, Arias AM. 1998. Different spatial and temporal interactions between Notch, wingless, and vestigial specify proximal and distal pattern elements of the wing in Drosophila. *Dev Biol* **194**: 196–212.

Lewis EB. 1978. A gene complex controlling segmentation in Drosophila. *Genes, Dev Cancer*

Life Work Edward B Lewis 229–242.

Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, et al. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* **47**: W636–W641.

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

Mann RS, Affolter M. 1998. Hox proteins meet more partners. *Curr Opin Genet Dev* **8**: 423–429.

McGinnis W, Krumlauf R. 1992. Homeobox genes and axial patterning. *Cell* **68**: 283–302.

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

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

Neumann CJ, Cohen SM. 1996. A hierarchy of cross-regulation involving Notch, wingless, vestigial and cut organizes the dorsal/ventral axis of the Drosophila wing. *Development* **122**: 3477–3485.

Nguyen NTT, Contreras-Moreira B, Castro-Mondragon JA, Santana-Garcia W, Ossio R, Robles-Espinoza CD, Bahin M, Collombet S, Vincens P, Thieffry D, et al. 2018. RSAT 2018: Regulatory sequence analysis tools 20th anniversary. *Nucleic Acids Res* **46**: W209–W214.

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.

Paul R, Giraud G, Domsch K, Duffraisse M, Marmigère F, Khan S, Vanderperre S, Lohmann I, Stoks R, Shashidhara LS, et al. 2021. Hox dosage contributes to flight appendage morphology in Drosophila. *Nat Commun* **12**. http://dx.doi.org/10.1038/s41467-021-23293-8.

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

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke

T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**: W160–W165.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29**: 24–26.

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

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

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

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189–1191.

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.

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

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

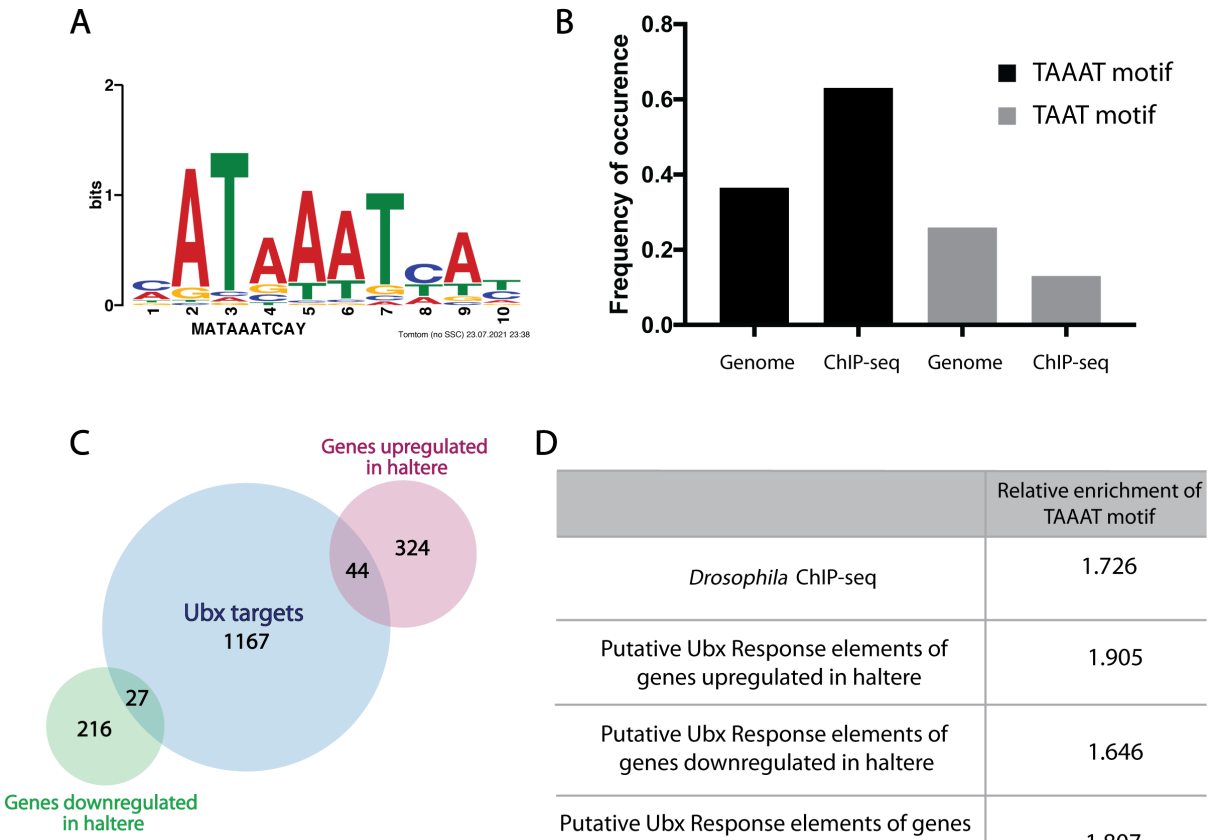
Williams JA, Bell JB, Carroll SB. 1991. Control of Drosophila wing and haltere development by the nuclear vestigial gene product. *Genes Dev* **5**: 2481–2495.

Yu G, Wang LG, He QY. 2015. ChIP seeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**: 2382–2383.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**.

Figure and Figure legends

Figure 1: A motif containing a TAAAT core sequence is enriched in the enhancers of putative direct targets of Ubx in *Drosophila* halteres



	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

Legend Fig. 1:

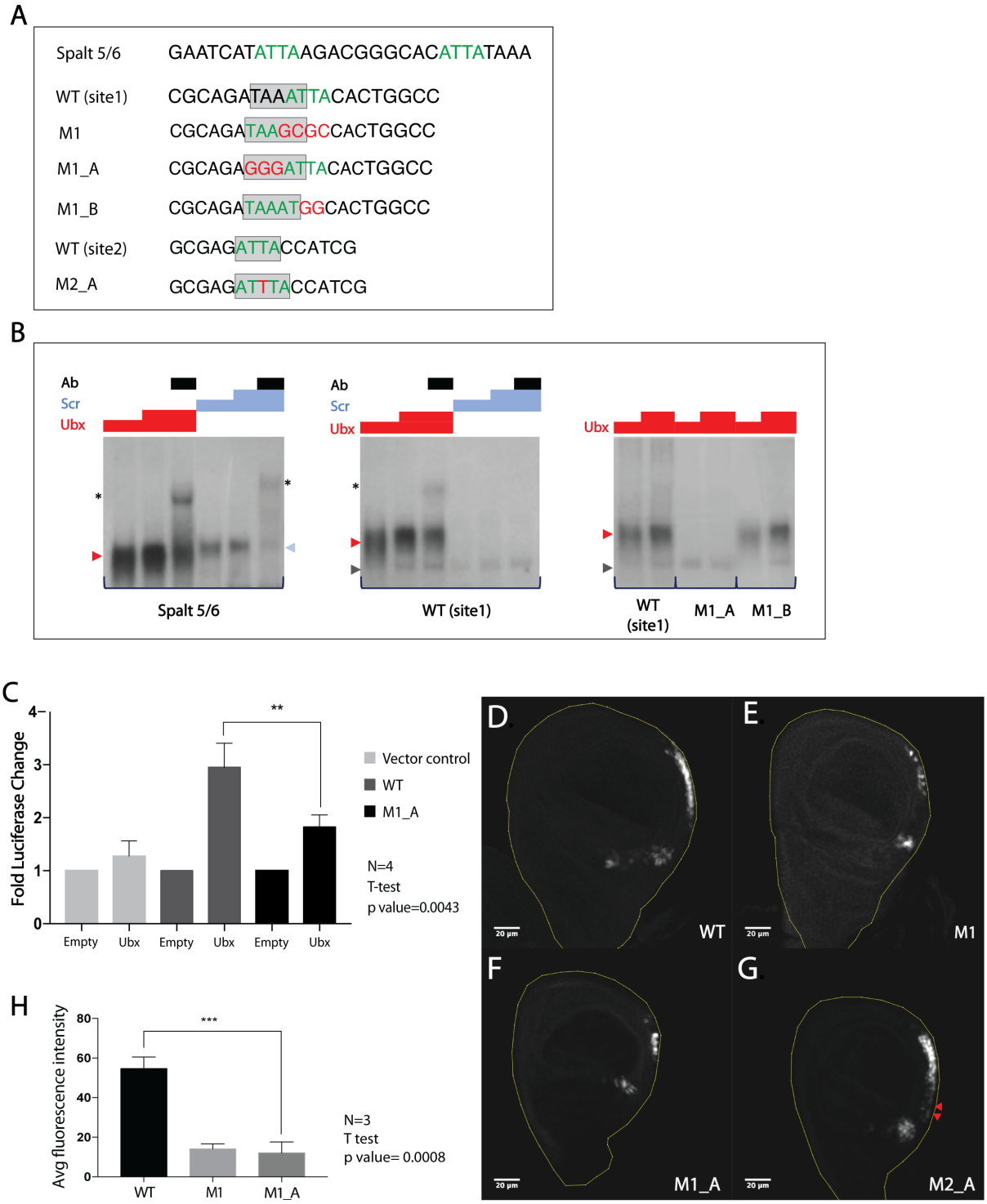
A) A motif containing TAAAT core sequence (hereafter referred to as TAAAT motif) is enriched by Ubx on denovo motif analysis of Ubx pulled down sequences from *Drosophila* halteres.

B) The frequency of the TAAAT motif in Ubx pulled down sequences is 1.7 fold more than the frequency of the motif in the entire genome. The motif containing TAAT core sequence (hereafter referred to as TAAT motif) was not enriched in Ubx pulled down sequences.

C) Comparison of Ubx direct targets to differentially expressed genes between the wing and haltere identified 27 targets that are downregulated and 44 targets that are upregulated by Ubx in the haltere.

D) Table showing the relative enrichment of the TAAAT and TAAT motifs in *Drosophila* ChIP-seq and Ubx Response elements of genes that are upregulated, downregulated or equally expressed between the wing and haltere imaginal discs in *Drosophila*

Figure 2: The TAAAT motif in the *CG13222* enhancer is critical for its regulation by Ubx



Legend Fig. 2:

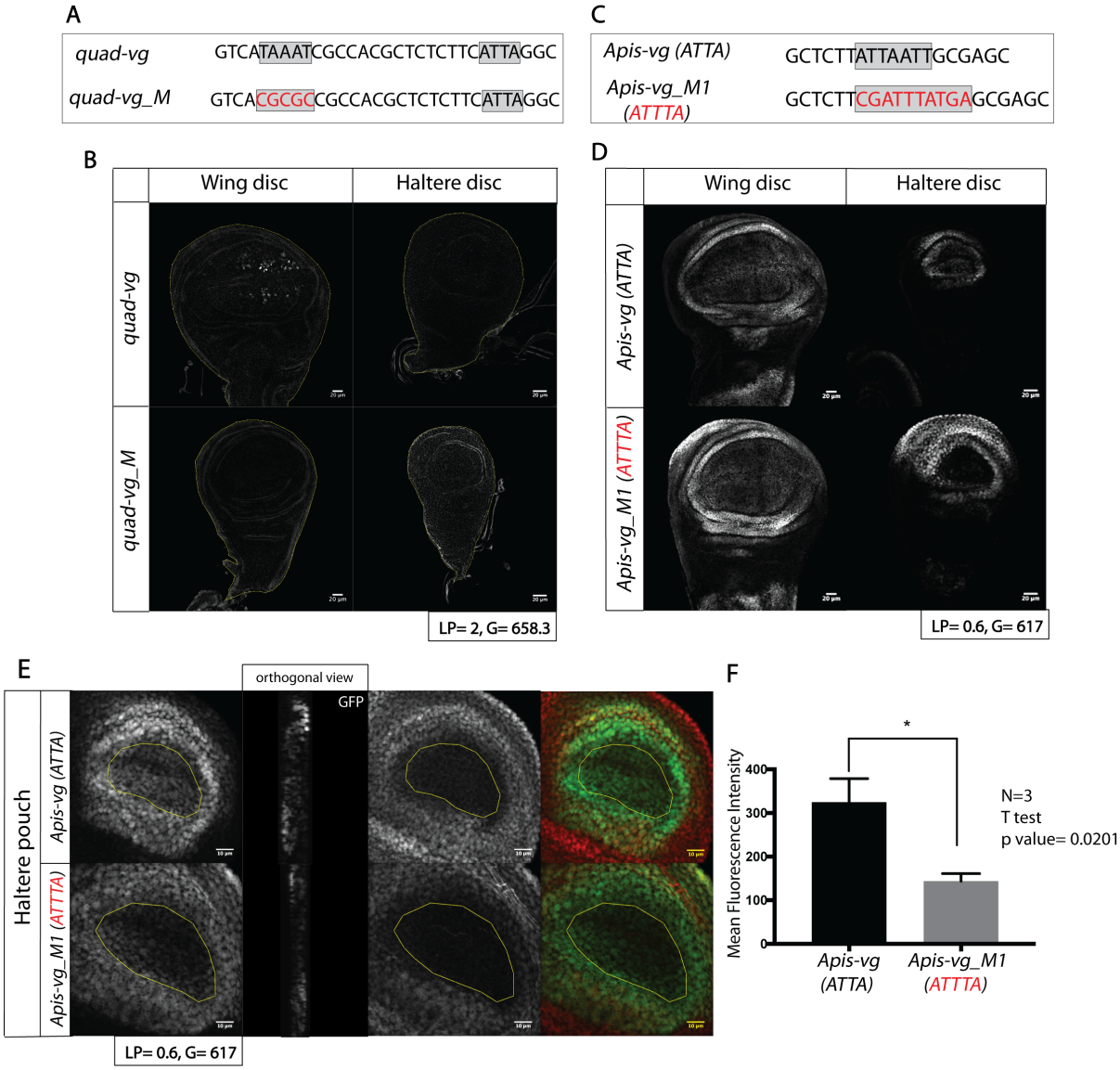
A) Sequence of part of *sal* enhancer (Spalt 5/6), *CG13222* enhancer (WT (site1)) and mutations of the *CG13222* enhancer (M1_A, M1_B) used as probes for EMSA. The wild type *CG13222* enhancer has two Ubx binding sites (shown in green) termed here as WT (site1) and WT (site2). Mutations were generated to specifically modify either the TAAT (M1_B, M2_A) or the TAAAT (M1_A) motifs while keeping the other one intact in site1 of the *CG13222* enhancer. For Luciferase and Transgenic assays, the entire 512bp spanning the *CG13222* edge enhancer was used.

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).

C) The *CG13222* enhancer drives Luciferase activity in a S2 cell culture system in presence of Ubx. Significant loss of reporter activity is observed in the mutant enhancer (M1_A) where the TAAAT motif is affected. For statistical analysis, t-test was performed (two-tailed).

D-G). The *CG13222* enhancer drives reporter expression in the posterior edge of the third instar haltere imaginal discs (D). The M1 mutant where both the TAAT and TAAAT motifs are affected show a marked reduction in reporter expression (E). The M1_A mutant carrying mutations in the TAAAT site of the enhancer has a similar phenotype as the M1 mutant (F) and shows a significant reduction in GFP expression as seen from calculation of average fluorescence intensity (H). Mutation of TAAT to TAAAT in site2 results in higher level of reporter expression along with ectopic expression in some domains (shown by red arrows) (G). For statistical analysis, t-test was performed (two-tailed).

Figure 3: The TAAAT site is sufficient to bring the orthologous enhancer of the *vestigial* gene from *Apis* under Ubx control in a transgenic assay



Legend Fig. 3:

A) Sequence of part of the 805bp enhancer of the *vestigial* gene in *Drosophila (quad-vg)* containing both TAAT and TAAAT motifs and its mutant enhancer (*quad-vg_M*).

B) The *quad-vg* drives expression of the reporter gene GFP in the pouch of wing imaginal discs but is repressed by Ubx in the haltere. Mutations in the TAAAT motif lead to complete loss of reporter gene expression from both wing and haltere imaginal discs. The laser power (LP) and gain (G) used for imaging GFP fluorescence is indicated.

C) Sequence of part of the 575bp enhancer of the *vestigial* gene in *Apis (Apis-vg)*, showing the presence of a TAAT motif the mutant enhancer (*Apis-vg_M1*), wherein the TAAT motif is changed to a TAAAT motif.

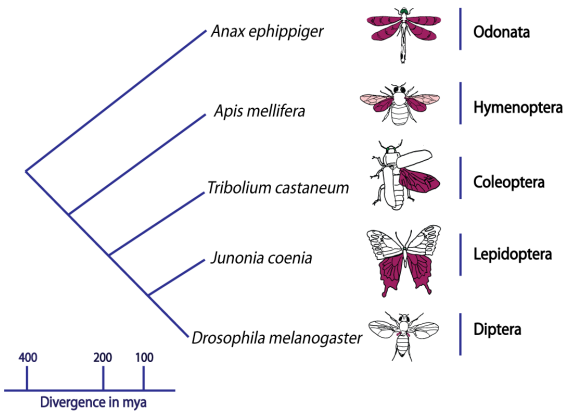
D) The *Apis-vg* enhancer drives expression of GFP in the pouch region in wing imaginal discs as well as haltere imaginal discs. The mutant enhancer (*Apis-vg_M1*) with introduced TAAAT motif shows reduced levels of GFP expression in the haltere pouch, while its expression in wing discs is not altered compared to the wild type enhancer.

E) Magnified images of the haltere pouch of *Drosophila* transgenics expressing GFP under *Apis-vg* and *Apis-vg_M1* enhancers. Note much reduced GFP levels 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 WT and the mutant enhancers. Hth staining, which is hinge-specific is used to demarcate the pouch region.

F) Quantification of average fluorescence intensity in the haltere pouch for the *Apis-vg* and *Apis-vg_M1* transgenics. For statistical analysis, t-test was performed (two-tailed).

Figure4: The TAAAT motif is not enriched in the enhancers of targets of Ubx in *Apis*

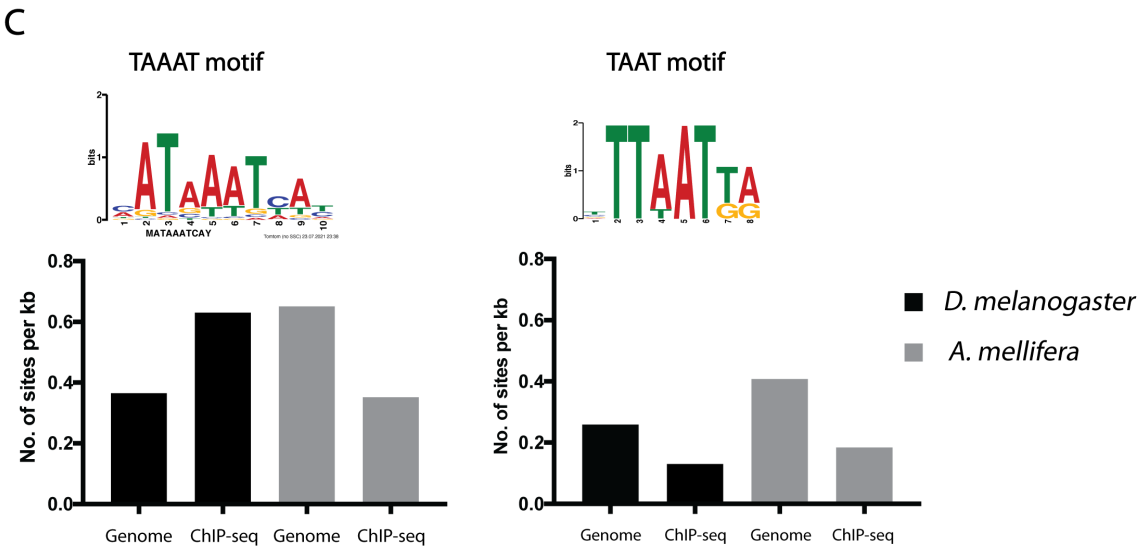
A



B

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

C



Legend Fig. 4:

A) While Ubx is required for haltere development in Dipterans such as *Drosophila melanogaster*, differences in hindwing morphology in the third thoracic segment of diverse insects is not correlated to the mere presence of Ubx.

B) De-novo motif analysis of Ubx ChIP-seq in *Apis* hindwings and *Drosophila* halteres reveal enrichment of the TAAAT motif specifically in *Drosophila* but not in *Apis*. Binding sites for transcription factors like Trl, CTCF and Brk are, however, enriched in both datasets.

C) FIMO analysis of the TAAAT motif reveals a 1.7fold enrichment of sites in *Drosophila* ChIP-seq sequences as compared to the entire genome. No such enrichment was found for the *Apis* ChIP-seq datasets. The TAAT motif was not enriched in either of the datasets.