Cis-regulatory modes of *Ultrabithorax* inactivation in butterfly forewings

Amruta Tendolkar[1], Anyi Mazo-Vargas[1], Luca Livraghi[1], Joseph J. Hanly[1,2], Kelsey C. Van Horne[1], Lawrence E. Gilbert[3], Arnaud Martin[1,*]

[1]: Department of Biological Sciences, The George Washington University, Washington, DC, United States

[2]: Smithsonian Tropical Research Institute, Panama

[3]: Department of Integrative Biology, University of Texas – Austin, Austin, TX, United States

*: Corresponding author:

Arnaud Martin, Department of Biological Sciences, The George Washington University, 800 22nd Street NW, Washington, DC 20052. Email: arnaud@gwu.edu

ABSTRACT

Hox gene clusters encode transcription factors that drive regional specialization during animal development: *e.g.* the Hox factor Ubx is expressed in the insect metathoracic (T3) wing appendages and differentiates them from T2 mesothoracic identities. *Hox* transcriptional regulation requires silencing activities that prevent spurious activation and regulatory crosstalks in the wrong tissues, but this has seldom been studied in insects other than *Drosophila*, which shows a derived *Hox* dislocation into two genomic clusters that disjoined *Antennapedia* (*Antp*) and *Ultrabithorax* (*Ubx*). Here we investigated how *Ubx* is restricted to the hindwing in butterflies, amidst a contiguous *Hox* cluster. By analysing Hi-C and ATAC-seq data in the butterfly *Junonia coenia*, we show that a Topologically Associated Domain (TAD) maintains a hindwing-enriched profile of chromatin opening around *Ubx*. This TAD is bordered by a Boundary Element (BE) that separates it from a region of joined wing activity around the *Antp* locus. CRISPR mutational perturbation of this BE releases ectopic *Ubx* expression in forewings, inducing homeotic clones with hindwing identities. Further mutational interrogation of two non-coding RNA encoding regions and one putative *cis*-regulatory module within the *Ubx* TAD cause rare homeotic transformations in both directions, indicating the presence of both activating and repressing chromatin features. We also describe a series of spontaneous forewing homeotic phenotypes obtained in *Heliconius* butterflies, and discuss their possible mutational basis. By leveraging the extensive wing specialization found in butterflies, our initial exploration of *Ubx* regulation demonstrates the existence of silencing and insulating sequences that prevent its spurious expression in forewings.

INTRODUCTION

Hox genes are key specifiers of antero-posterior regional identity in animals, and thus require robust regulatory mechanisms that confine their expression to well-delimited sections of the body. Their genomic arrangement into *Hox* gene clusters has provided a rich template for the study of gene regulation, with mechanisms including chromatin silencing and opening, 3D conformational changes, and non-coding RNAs (Mallo and Alonso 2013). However, this rich body of work has been almost exclusively performed in mice and fruit flies. In order to decipher how diverse body plans and morphologies evolved, we must start assessing the mechanisms of Hox gene regulation in a wider range of organisms.

The *Ultrabithorax* (*Ubx*) gene encodes a Hox family transcription factor involved in the specification of segment identities in arthropods (Hughes and Kaufman 2002; Heffer and Pick 2013). In insects, the conserved expression of *Ubx* in the metathoracic (T3) segment is required for their differentiation from Ubx-free tissues in the mesothorax (T2) , and has been a key factor for the specialization of metathoracic serial appendages including T3 legs (Mahfooz et al. 2007; Refki et al. 2014; Tomoyasu 2017; Feng et al. 2022; Buffry et al. 2023) and hindwings or their derivatives (Tomoyasu 2017; Loker et al. 2021). The mechanisms of *Ubx* segment-specific expression have been intensively studied in *D. melanogaster* (Mallo and Alonso 2013; Hajirnis and Mishra 2021), where *Hox* genes are separated into two genomic loci, the Antennapedia (ANT-C, *Antp*) and Bithorax clusters (BX-C). In short, the BX-C complex that includes *Ubx*, *abdominal-A* (*abd-A*), and *Abdominal-B* (*Abd-B*) is compartmentalized into nine chromosomal domains that determine the parasegmental expression boundaries of these three genes (Maeda and Karch 2015). Each boundary is primarily enforced by insulators that separate Topologically Associating Domains (TADs) of open-chromatin, while also allowing interactions of enhancers with their target promoters (Postika et al. 2018; Srinivasan and Mishra 2020). The BX-C locus also includes non-coding RNAs, some of which are processed into miRNAs known to repress *abd-A* and *Abd-B* (Garaulet and Lai 2015). *Fub-1/bxd* long non-coding RNAs (lncRNAs) situated 5' of *Ubx* are thought to participate in *Ubx* regulation in the PS5 (posterior T3 to anterior A1) parasegment (Ibragimov et al. 2022). An intronic lncRNA dubbed *lncRNA:PS4* is expressed in the PS4 parasegment (posterior T2 - anterior T3), and appears to stabilize *Ubx* in this region in mutant contexts (Hermann et al. 2022). Little is known about how *Hox* genes are regulated outside of flies, where they co-localize into a single *Hox* cluster, and where *Antp* and *Ubx* thus occur in contiguous positions (Gaunt 2022; Mulhair and Holland 2022). A few *Hox*-related miRNAs are evolutionarily conserved across the locus in arthropods (Pace et al. 2016), and an early study in *Tribolium* characterized the embryonic expression of a *Hox* cluster non-coding transcripts (Shippy et al. 2008).

These knowledge gaps lead us to consider the use of butterflies and moths (Lepidoptera) as alternative model systems for the study of *Ubx* function and regulation. Lepidopteran forewings and hindwings are functionally and morphologically differentiated, and CRISPR mosaic knock-outs (mKOs) showed that *Ubx* is necessary for the specification of hindwing color patterns, shape, and venation (Tendolkar et al. 2021). In three species of nymphalid butterflies (*Heliconius erato*, *Junonia coenia*, and *Bicyclus anynana*), CRISPR-mediated loss-of-function of *Ubx* induces regional-specific homeotic transformations of hindwing patterns into their forewing counterpart (Matsuoka and Monteiro 2018; Tendolkar et al. 2021), reminiscent of homeotic aberrations that are sporadically observed in butterfly wings (Sibatani 1983; Nijhout and Rountree 1995). The ectopic activation of *Ubx* into the pupal forewing results in the gain of hindwing features, suggesting Ubx is sufficient to drive T3-like identity when

expressed in T2 (Lewis et al. 1999; Tong et al. 2014). Besides its roles in adult wing differentiation, *Ubx* is also known to repress thoracic leg identity in transient embryonic appendages of the first abdominal segment, called pleuropods (Zheng et al. 1999; Masumoto et al. 2009; Tong et al. 2017; Tendolkar et al. 2021; Matsuoka et al. 2022).

The general organization of Hox gene clusters has been well described in Lepidoptera, but their regulation has been seldom studied. Lepidopteran genomes have accumulated divergent *Hox3* copies, named *Shox* genes, that are required during early embryonic development but do not appear to play homeotic functions (Ferguson et al. 2014; Livraghi 2017; Mulhair et al. 2022). An lncRNA and two miRNAs were identified in the intergenic region between *abd-A* and *Abd-B* in the silkworm (Wang et al. 2017, 2019). In butterfly wings, the regulation of *Ubx* shows strong patterns of segment-specific regulation at two levels. First, the *Ubx* transcript has been consistently identified as the most differentially expressed mRNA between the two wing sets (Hanly et al. 2019; Wang et al. 2022). Second, comparison of ATAC-seq signals reveal that forewing vs. hindwing have identical open-chromatin profiles during development across the genome, except at the *Ubx* gene itself (Lewis and Reed 2018; van der Burg et al. 2019). Thus, the ability of the *Ubx* locus to be robustly activated in hindwings and repressed in forewings is likely driving most subsequent differences between these tissues. In this study, we provide an initial assessment of the regulation of the *Ubx* locus during butterfly wing development. To do this, we leverage genomic resources and CRISPR mutagenesis with a focus on two laboratory species belonging to the Nymphalinae sub-family, *J. coenia* and *Vanessa cardui* (Livraghi et al. 2017; Martin et al. 2020; van der Burg et al. 2020; Mazo-Vargas et al. 2022). We identify putative regulatory regions likely involved in the repression and activation of *Ubx* expression, and discuss the potential mechanisms restricting it to hindwings. Finally, we describe a collection of spontaneous wing homeotic mutants in *Heliconius spp.* and elaborate on the categories of mutations that could underlie these phenotypes by misregulating *Ubx*.

RESULTS

Gene expression at the *Ubx* locus during wing development

We provide annotations of the *Ubx* genomic region in four Nymphalinae butterflies (**Fig. 1A**). These feature existing genomic resources for our model species *J. coenia* and *V. cardui* (van der Burg et al. 2020; Lohse et al. 2021b; Zhang et al. 2021), as well as for *Aglais* (*Nymphalis*) *io* (Lohse et al. 2021a). The publicly available annotations for these three species include evidence from developmental transcriptomes, and we added to this set a manual annotation of the *Ubx* locus from the oak leaf butterfly *Kallima inachus*, for which forewing *vs.* hindwing transcriptomes have been sequenced across a replicated developmental time series (Yang et al. 2020; Wang et al. 2022).

All Nymphalinae show a similar organization of the region spanning *Ubx*. Interestingly, the first intron of *Ubx* encodes a long non-coding RNA in opposite orientation to *Ubx*, that we dub here *lncRNA:Ubx-IT1* (abbr. *Ubx-IT1*), based on the recommended nomenclature (Seal et al. 2022). Orthologous versions of *Ubx-IT1* are detected in most NCBI RefSeq genome annotations throughout Lepidoptera (*e.g.* the ncRNA *NCBI:XR_960726* in *Plutella xylostella*), implying it is a conserved feature of the *Ubx* locus in this insect order. Finally, both annotations from *V. cardui, A. io,* and *J. coenia* show a long intergenic non-coding transcript starting in antisense orientation about 10-15 kb 5' of *Ubx*, that we dub here *lincRNA:Ubx-AS5'* (abbr. *Ubx-AS5'*). This transcript was neither detected in *K. inachus* transcriptomes nor in RNA datasets outside of the Nymphalinae sub-family, and could be specific to this lineage.

Next we reanalyzed the *K. inachus* wing transcriptomes (Wang et al. 2022), and profiled the expression of *Ubx* region transcripts during wing development (**Fig. 1B**). As expected from previous studies (Hanly et al. 2019; Paul et al. 2021; Merabet and Carnesecchi 2022; Wang et al. 2022), *Ubx* showed a strong expression bias in hindwings, spanning the larval imaginal disks to the intermediate pupal stage. Of note, Ubx is confined to the peripodial membranes of insect T2 wing appendages (Weatherbee et al. 1998, 1999; Prasad et al. 2016), which may explain residual detection in some of the forewing samples here. *Ubx-IT1* was significantly enriched in hindwings compared to forewings, albeit at ~100-fold lower count levels than *Ubx* in the same samples. The *Hox* gene *Antp* showed a minor forewing enrichment, confirming that while *Ubx* expression is robustly repressed in T2 forewing tissues, *Antp* expression is permitted in both T2 and T3 appendages (Matsuoka and Monteiro 2021, 2022; Paul et al. 2021). Expression of *abd-A* was undetected in most wing samples.

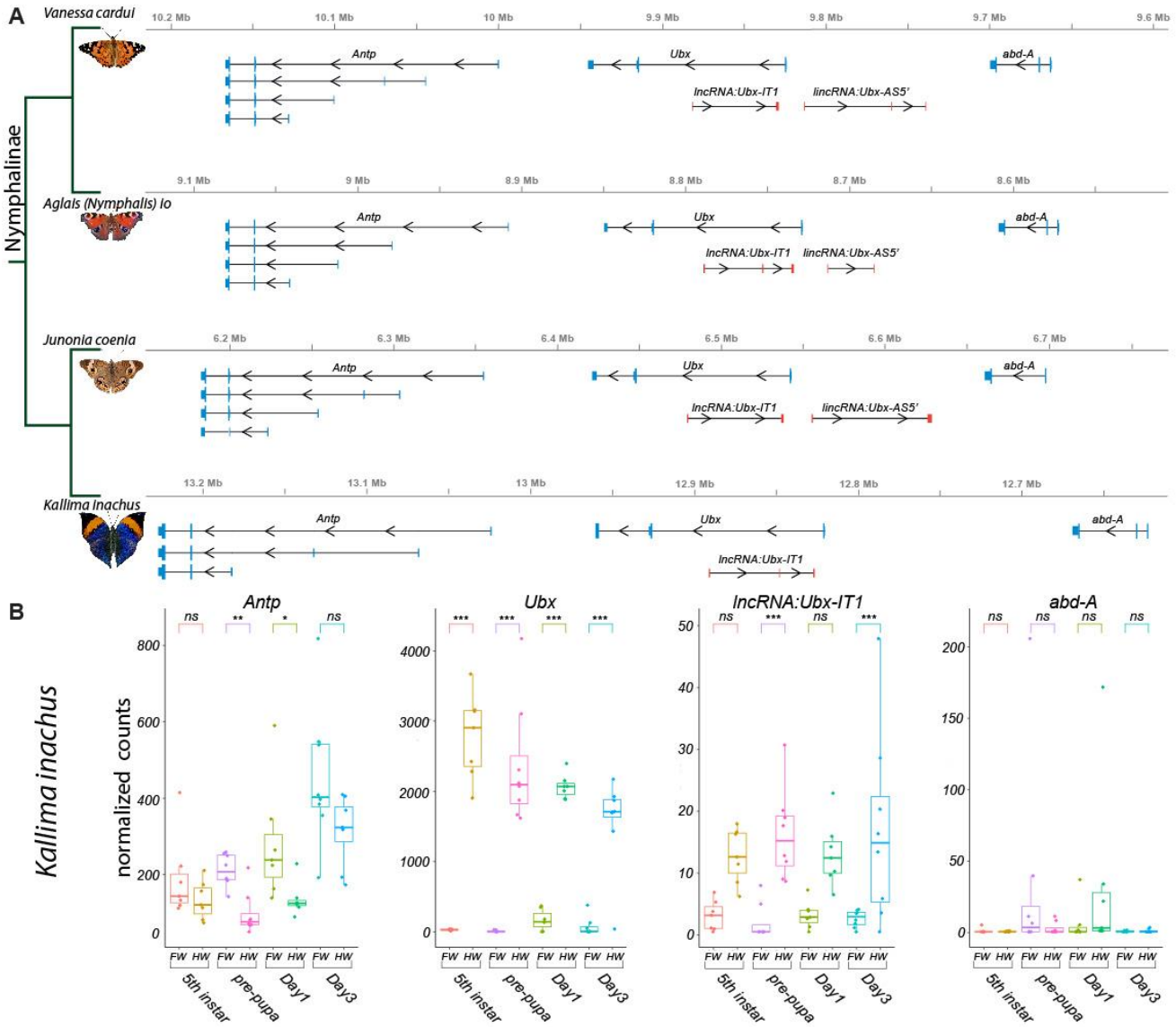


Figure 1. Annotation of the *Ubx* genomic interval in four butterflies of the Nymphalinae sub-family. (A) Genomic intervals spanning *Antp*, *Ubx*, and *abd-A*, featuring published transcript annotations from NCBI Reference Genomes of *V. cardui* and *A. io*, and manual re-annotations of the *J. coenia* and *K. inachus* genomes using published RNAseq dataset (see Methods). Exons are shown with coding (thick) and non-coding (thin) sections. No *lincRNA:Ubx-AS5'* transcripts were detected in *K. inachus*. **(B)** Expression profiling of transcripts of the *Ubx* region in *K. inachus*, based on a reanalysis of published wing RNA-seq transcriptomes (Wang et al. 2022). Expression levels are plotted as DESeq2 normalized counts plots. Pairwise Wald tests adjusted for multiple test correction each assess differential expression between forewings and hindwings. ns : non-significant ; * : $p < 0.05$; ** : $p < 0.01$; *** : $p < 0.001$.

Chromatin 3D conformation reveals a Boundary Element between *Antp* and *Ubx*

Genome-wide Hi-C sequencing can be used to generate heatmaps that capture the conformation of 3D chromatin in tissues, and has been used extensively to study *Drosophila Hox* cluster organisation into TADs that prevent regulatory crosstalk between adjacent genes (Ibragimov et al. 2022; Moniot-Perron et al. 2023). Here we used Hi-C to assess the 3D chromatin architecture of the *Hox* cluster interval in the butterfly *J. coenia*, using existing datasets that were generated from fifth instar larval forewings (van der Burg et al. 2020; Mazo-Vargas et al. 2022). In larval forewings, the *Hox* chromatin conformation landscape consists of three well-delimited TADs, the first one spanning *proboscipedia* (*pb*) to *Sex comb reduced* (*Scr*), the second one around *Antp*, and the third one *Ubx*, *abd-A, and Abd-B* (**Figs. 2 and 3A**). A Boundary Element (BE), was robustly called (see Methods) at the region separating the *Antp* and *Ubx* TADs, situated in the *Ubx* last intron. Because TAD boundary prediction has a coarse resolution, we arbitrarily define the candidate BE region as a 15-kb interval centered in the *Ubx* last intron, and dub it *Antp-Ubx_BE*. A binding motif analysis identified 4 CTCF binding sites in a 1-kb interval within *Antp-Ubx_BE*, two of which were found in a tightly linked, convergent orientation (**Fig. S1**), which is consistent with TAD insulating role in mediating chromatin loop-extrusion (Guo et al. 2015). This concordance between Hi-C profiling and CTCF motif prediction thus indicates that *Antp-Ubx_BE* region functions as an insulator between regulatory domains of *Antp* and *Ubx*.

Differential forewing vs. hindwing chromatin-opening across the *Antp-Ubx* interval

In flies, the *Ubx/abd-A* section is organized into regulatory domains that display differential activities across the antero-posterior axis, following what has been called the open-for-business model (Maeda and Karch 2015; Gaunt 2022). Here we tested if this pattern extends to butterfly species with a contiguous *Hox* cluster. To do this we used ATAC-seq datasets from *J. coenia* forewing (T2), hindwing (T3), and whole-head tissues sampled across fifth instar larval and early pupal stages, similarly to previous studies (van der Burg et al. 2020; Mazo-Vargas et al. 2022; Van Belleghem et al. 2023). These data reveal that chromatin opening along the *Antp/Ubx/abd-A* interval is partitioned into several regions showing a transition of T2 to T3 activity (**Fig. 2B**). From the anterior to posterior *Hox* colinear order (*i.e.* from *Antp* towards *abd-A*), chromatin-opening forms a block of forewing-enriched activity close to *Antp* and its 5' region, to a block of activity in both forewings and hindwings that stops at the *Antp-Ubx_BE*. This region is consistent with the fact that *Antp* is expressed in both wing pairs (**Fig. 1B**). From *Antp-Ubx_BE,* the interval including *Ubx* and a large upstream region is strongly enriched for hindwing opening, consistently with previous studies that found it to be the only genomic region showing this pattern (Lewis and Reed 2018; van der Burg et al. 2019). Last, the region surrounding *abd-A* is devoid of differntial open-chromatin activity between forewings and hindwings, in accordance with the exclusion of its expression from thoracic segments (Warren et al. 1994; Tong et al. 2014)

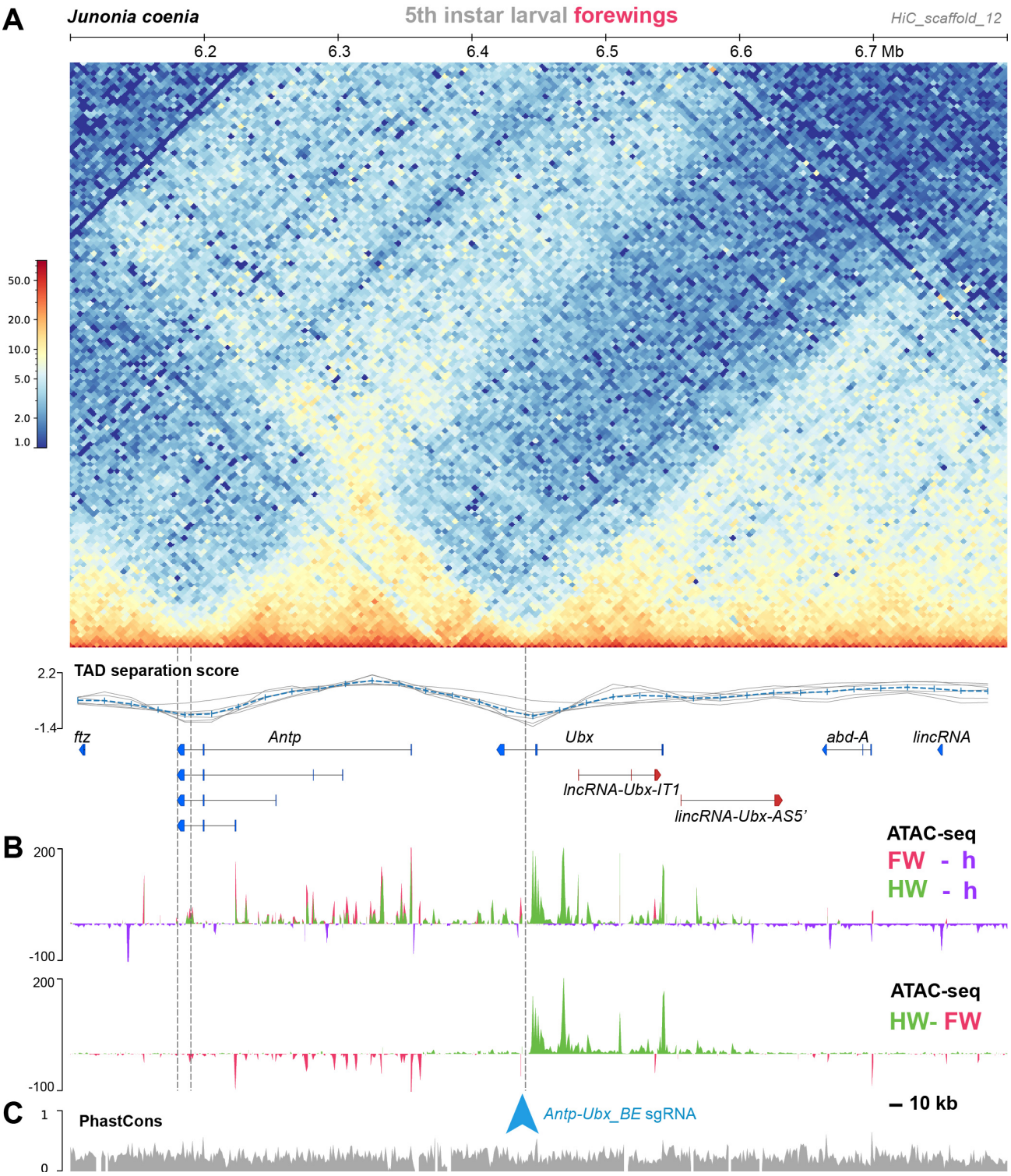


Figure 2. A region of hindwing-specific chromatin-opening is bordered by a TAD BE in the last intron of *Ubx*. **(A)** Hi-C contact heatmap in fifth instar forewings of *J. coenia* and TAD separation scores around *Ubx*. A TAD boundary element (*Antp-Ubx_BE*) is inferred in the last intron of *Ubx* (vertical dotted line). **(B)** Differential ATAC-seq profiles, re-analyzed from a previous dataset (Mazo-Vargas et al. 2022). Top : open-chromatin profiles of forewings (FW, magenta), and hindwings (HW, green), respectively subtracted from larval head signal (purple, negative when wing signals at background-level). Bottom : subtractive ATAC-seq profile (HW-FW) revealing hindwing-enriched chromatin in the *Ubx* locus. *Antp-Ubx_BE* is in the vicinity of an isolated region of forewing-enriched opening (blue arrowhead). **(C)** *PhastCons* genomic alignment scores, with overall alignability suggesting minimal structural variation across this interval in Lepidoptera and Trichoptera.

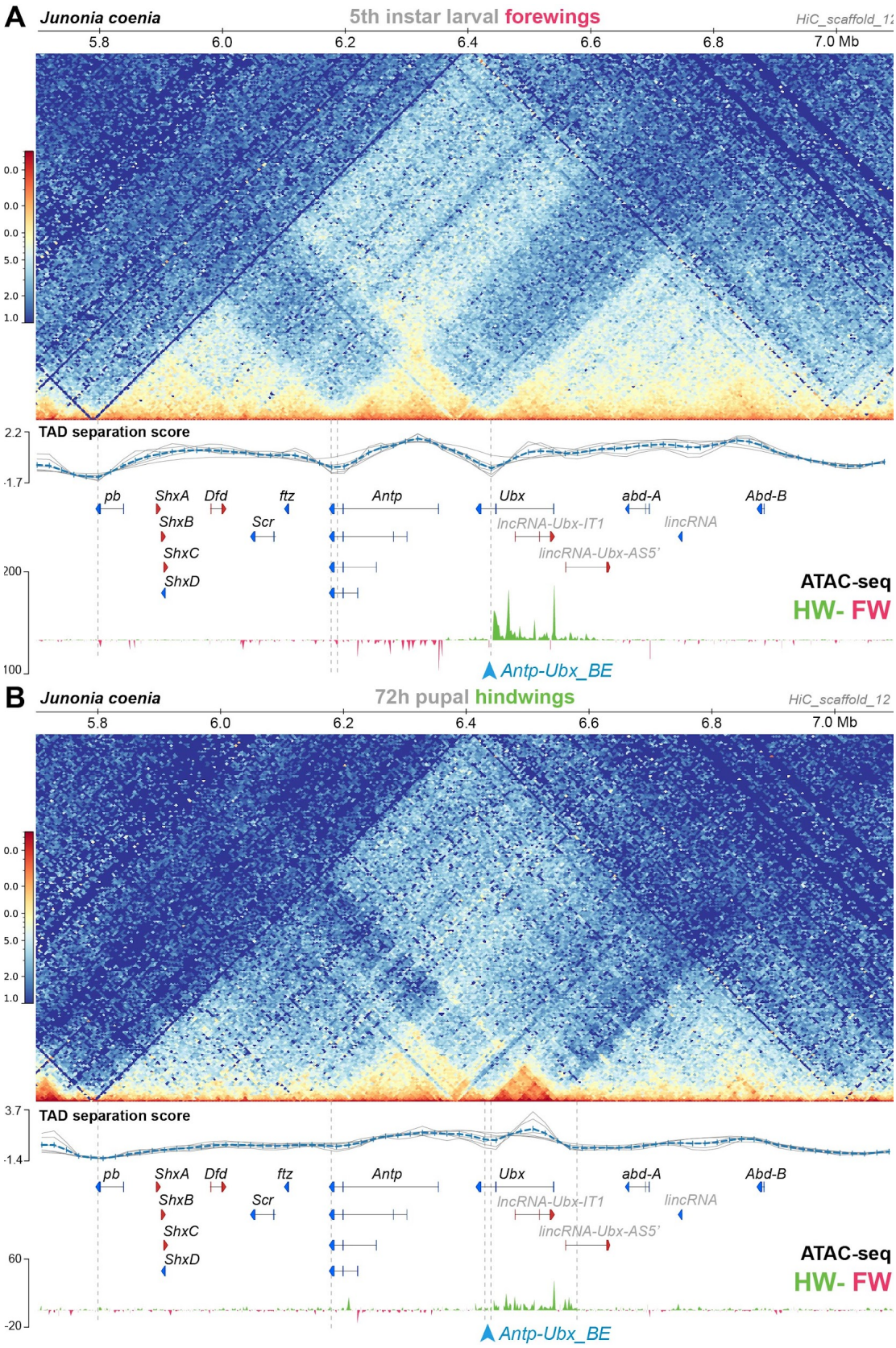


Figure 3. Hindwing-enriched chromatin-opening around *Ubx*, and the *Antp-Ubx_BE* boundary, are both maintained in mid-pupal hindwings. (A) Hi-C heatmap in *J. coenia* fifth instar larval forewings, and subtractive ATAC-seq profiles at this stage (hindwing-forewing), as expanded from Fig. 2 across the *Hox* cluster. **(B)** Hi-C heatmap in *J. coenia* mid-pupal hindwings, and subtractive ATAC-seq profiles at this stage (forewing-hindwing). Inferred TAD boundaries are shown as vertical dotted lines. Blue arrowhead : position of the *Antp-Ubx_BE* sgRNA.

Comparison of 3D conformation and open-chromatin profiles between larval forewings and mid-pupal hindwings

The Hi-C dataset analyzed above was prepared from larval forewings, and forewings do not express Ubx (**Fig. 1B**). Next, we repeated our analysis on a Hi-C dataset generated in pupal hindwings instead (van der Burg et al. 2020), *i.e.* in a later-stage tissue expressing *Ubx*. We found two main differences in this contact landscape compared to the larval forewing (**Fig. 3**). First, the TAD spanning from *proboscipedia* (*pb*) to *fushi-tarazu* (*ftz*) faded in intensity, while in contrast, the TAD around *Antp* remained strongly organized. Second, *Ubx* lost its physical association to the *abd-A* and *Abd-B* domains, and gained a TAD boundary situated in the *Ubx-AS5'* intron. It is difficult to disentangle effects from staging (larval vs. pupal) and tissues (forewing vs. hindwing) in this comparison. Specifically, these differences we observed may be due to chromatin remodeling between stages, as somewhat expected during metamorphosis (Gutierrez-Perez et al. 2019). Alternatively, it is also possible hindwing development requires insulate *Ubx* from more posterior enhancers. These issues will require further investigation, for instance using profiling of histone marks, with pairwise forewing-hindwing comparison at single stages. Nonetheless the later hindwing sample showed a maintenance of *Antp-Ubx* separation. First, while *Ubx* formed a smaller TAD spanning its coding exons 1-2, this region conserved a domain of hindwing-enriched open-chromatin. Second, boundary prediction called two possible, tightly linked TAD limits in the *Antp-Ubx_BE* region, showing that the last intron of *Ubx* still acts as an insulating region. In conclusion, our preliminary comparison of *Hox* 3D conformation indicates that the *Antp-Ubx_BE* is relatively stable across two stages and wing serial homologs.

Mutagenic perturbation of *Antp-Ubx_BE* results in forewing homeosis

Next, we reasoned that the forewing-enriched ATAC-seq peak present in the inferred boundary interval (**Fig. 4A**) might mediate the binding of insulator proteins (Savitsky et al. 2016; Stadler et al. 2017), or act as a transcriptional silencer (Segert et al. 2021). Several genomic features support the former hypothesis. First, the only forewing-enriched ATAC-seq peak across a 150-kb region (spanning the *Ubx* gene and the *Antp-Ubx* intergenic region), coincides with the midpoint between the two tentative *hicFindTADs* boundary predictions inferred from HiC data (**Fig. 2B**). Second, during motif scans conducted across that 150-kb region we found 8 predicted binding-sites for the *Drosophila* CCCTC-Binding Factor (CTCF) clustered in a 5-kb region around the differentially accessible region, and none elsewhere in the last Ubx intron (**Fig. 4A**), suggesting the forewing-enriched ATAC-seq peak may function as a transcriptional insulator (Gambetta and Furlong 2018; Postika et al. 2018; Kyrchanova et al. 2020; Kaushal et al. 2022). Last, the two candidate CTCF binding motifs that are within the forewing-enriched ATAC-seq peak are also conserved across Lepidoptera and Trichoptera (**Fig. S1**), two lineages that diverged around 300 Mya (Kawahara et al. 2019; Thomas et al. 2020).

To test this hypothesis, we used CRISPR targeted mutagenesis to perturb *Antp-Ubx_BE* and assess its functionality, and designed a single sgRNA in a conserved sequence within the forewing-enriched ATAC-seq (**Fig. S1**). Remarkably, CRISPR mutagenesis of the *Antp-Ubx_BE* target induced G_0 mutants with homeotic transformations of their forewings into hindwings (**Figs. 4B-C and S2**), including identity shifts in patterns, venation, and wing shape. It is important to note that none of the resulting crispants showed hindwing effects. Thus, we can reasonably attribute forewing homeotic phenotypes to indel mutations restricted to the intronic region, without disruption of the *Ubx* transcript, as this would result in hindwing phenotypes (Matsuoka and Monteiro 2021; Tendolkar et al. 2021).

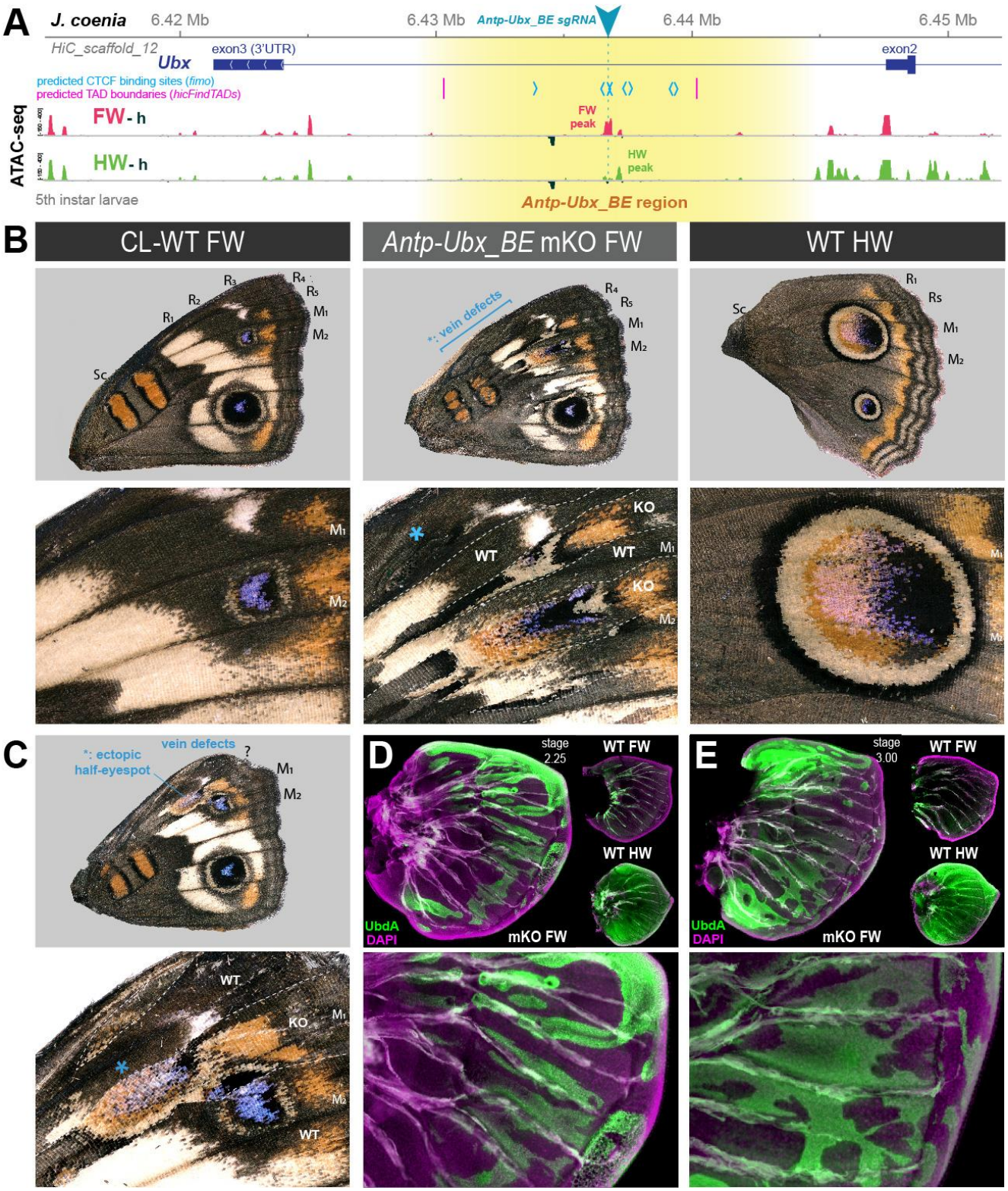


Figure 4. CRISPR perturbation of *Antp-Ubx_BE* results in FW→HW homeoses. (A) *Antp-Ubx_BE* sgRNA targeting (cyan triangle) of a FW-enriched ATAC-peak (magenta) within the *Ubx* last intron. **(B-C)** Two examples of *J. coenia Antp-Ubx_BE* crispants showing mosaic FW→HW homeoses, shown in dorsal views. CL-WT : contralateral, horizontally flipped images of forewings from the same individuals. WT HW : wild type hindwings from the same individual and mutant forewing side. Both individuals show disruption of their Radial veins (R_1-R_5 area). The specimen shown in C displays a partial, ectopic eyespot (asterisk). (**D-E**) Immunofluorescent detection of the UbdA epitope (green) in fifth instar wings disks of *Antp-Ubx_BE* crispants, revealing ectopic antigenicity in forewings. WT forewings of similar stage, and HW from the same crispant individuals, are shown for comparison as insets. Green autofluorescence was observed in tracheal tissues.

236 Homeotic clones are first visible in *Antp-Ubx_BE* crispants at the pupal stage, with streaks of
237 thinner cuticle, sometimes associated with local necrosis or with suture defects in the ventral midline, in
238 particular where leg and wing pouches adjoin (**Fig. S3**). Color pattern homeotic clones were salient at
239 the adult stage, with for example, clonal losses of the forewing specific white-band, and partial
240 acquisitions of the large M_1-M_2 hindwing eyespot. In one specimen, an ectopic, partial M_1-M_2 hindwing
241 eyespot appeared in the R_5-M_1 region, suggesting a perturbation of the eyespot induction process in
242 this wing. Nymphalid forewings have five radial veins (R_{1-5}), which provide sturdiness for flight (Wootton
243 1993), while hindwings have only two Radial veins. Forewing homeotic mutants showed mosaic venation
244 defects in the Radial vein area (**Fig. 4B**). Finally, higher expressivity mutant forewings were smaller and
245 rounder, reminiscent of hindwing shape.
246 Next, we dissected fifth instar larval wing disks from developing *Antp-Ubx_BE* crispants, and
247 monitored the expression of Ubd-A (Ubx and Abd-A epitopes), normally restricted to the hindwing and
248 only present in the forewing peripodial membrane (Weatherbee et al. 1999). Crispants showed forewing
249 clones with strong ectopic expression of Ubd-A (**Figs. 4D-E and S4**). This result supports the inference
250 that *Antp-Ubx_BE* forewing homeoses are due to the de-repression of *Ubx* in this tissue.
251
252 **Mutational interrogation of lncRNA-encoding regions at the *Ubx* locus**
253 We used CRISPR mutagenesis to test the function of the two lncRNA-encoding loci at the *Ubx* locus.
254 Mutagenesis of the *Ubx-IT1* first exon in *J. coenia*, and of the *Ubx-T1* promoter in *V. cardui*, both
255 resulted in crispants with small homeotic phenotypes in forewings and hindwings (**Figs. 5 and S5**). This
256 result contrasts with *Ubx* exon mKO experiments, which only generate hindwing phenotypes (Tendolkar
257 et al. 2021). Given the scarcity of *Ubx-IT1* crispants obtained (11 out of 236 adults), and the small size of
258 the homeotic clones within them, we infer the occasional phenotypes may be due to rare alleles. Thus,
259 rather than evidence of functionality of the *Ubx-IT1* transcript, the homeotic phenotypes may rather
260 reflect the effects of regulatory perturbation on *Ubx* itself, with some random mutations in this intronic
261 region resulting in hindwing *Ubx* loss-of-function, and some others triggering derepression in forewings.
262 Likewise, next we mutagenized the first exon of *Ubx-AS5'*, located upstream of the *Ubx* promoter, and
263 obtained twelve hindwing mutants and a single forewing mutant (**Fig. 6 and S6**). As with *Ubx-IT1*
264 CRISPR experiments, these results may be explained by regulatory disruption of *Ubx* transcription, with
265 a higher ratio of hindwing phenotypes compared to forewings linked to the proximity of the *Ubx*
266 promoter. Overall, we conclude that the mutational interrogation at these loci can result in dual loss
267 (hindwing) and gain (forewing) of *Ubx* function effects. Deciphering whether or when these effects
268 affected *Ubx* expression via local *cis*-regulatory modules, impairment of lncRNA transcripts, or larger
269 indels overlapping with *Ubx* exons, will require further study (see Discussion).

Table 1. CRISPR mutational interrogation experiments at putative *Ubx* regulatory regions

Species	sgRNA(s)	Inj. Embryos Ninj	L1 larvae Nhat	Pupae or *L5 larvae*	Adults Nadu	Crispants Nmut	Inj. time h AEL	Cas9:sgRNA ng/µL	Hatching Rate Nhat/Ninj	Crispant Rate Nmut/Ninj
J. coenia	*Antp-Ubx_BE*	59	50	50	44	6	2.5-3.5	500 : 250	84.7%	10.2%
		118	40	40	31	6	1.75-2.75	250 : 125	33.9%	5.1%
		89	44	*44*	*39* *	17	2.25-3.5	500 : 250	49.4%	19.1%
	Total	**266**	**90**	**134**	**115**	**29**			**33.8%**	**10.9%**
V. cardui	*IT1_sgRNA1*	204	67	50	50	2	1-3	250 : 125	32.8%	1.0%
		108	49	31	31	3	2-3	125 : 62.5	45.4%	2.8%
		145	60	39	39	2	2.25-3.5	500 : 250	41.4%	1.4%
	Total	**457**	**176**	**120**	**120**	**7**			**38.5%**	**1.5%**
J. coenia	*IT1_sgRNA2*	59	40	7	6	0	0.5-2.5	500 : 250	67.8%	0.0%
		124	112	112	110	4	2.25-4.75	500 : 250	90.3%	3.2%
	Total	**183**	**152**	**119**	**116**	**4**			**83.1%**	**2.2%**
V. cardui	*AS5_sgRNA1*	334	183	57	52	5	2-3	250 : 125	54.8%	1.5%
		122	87	2	2	0	2-4	500 : 250	71.3%	0.0%
	Total	**456**	**270**	**59**	**54**	**5**			**59.2%**	**1.1%**
J. coenia	*AS5_sgRNA1*	**309**	**181**	**181**	**181**	**8**	2-4.5	500 : 250	**58.6%**	**2.6%**
J. coenia	*Ubx11a2+3 +c5+6*	317	18	-	-	2	1-3	500 : 75 ea.	5.7%	0.6%
		203	35	0	0	0	1.5-3.5	500 : 75 ea.	17.2%	0.0%
	Total	**520**	**53**	**-**	**-**	**2**			**10.2%**	**0.4%**
V. cardui	*Ubx11a2+c5*	50	5	3	3	2	4-4.5	500 : 500	10.0%	4.0%
		151	29	6	5	2	2-2.75	500 :125:125	19.2%	1.3%
		361	18	13	16	6	0.5-2	500 :125:125	5.0%	1.7%
	Total	**562**	**52**	**22**	**24**	**10**			**9.3%**	**1.8%**
V. cardui	*Ubx11c5*	168	99	27	26	3	3.75-4.75	250 : 125	58.9%	1.8%
		62	22	9	9	2	0.5-0.75	500 : 250	35.5%	3.2%
		131	93	8	8	3	1.5-3	500 : 250	71.0%	2.3%
		114	63	20	20	6	3.5-4.5	500 : 250	55.3%	5.3%
	Total	**475**	**277**	**64**	**63**	**14**			**58.3%**	**2.9%**
V. cardui	*Ubx11b9*	32	18	6	5	1	1.25-2.25	500 : 250	56.3%	3.1%
		63	49	9	6	1	3.5-4.5	500 : 250	77.8%	1.6%
	Total	**95**	**67**	**15**	**11**	**2**			**70.5%**	**2.1%**
J. coenia	*Ubx11b9*	41	13	13	13	3	2.5-4	125 : 62.5	31.7%	7.3%
		48	21	14	14	1	2-3	250 : 125	43.8%	2.1%
	Total	**89**	**34**	**27**	**27**	**4**			**38.2%**	**4.5%**

* : upper estimate, includes 16 fifth instar larvae that were dissected for immunostainings, of which 7 were mutants (FW UbdA+), and 3 dissected mutant pupae.

Figure 5. Rare, dual homeoses obtained from CRISPR mutagenesis of the *lncRNA_Ubx-IT1* 5' region. **(A)** Genomic context of the sgRNA targets (here shown in *J. coenia*), in the promoter and first exon of the non-coding *Ubx-IT1* transcript. **(B-C)** Dorsal and ventral views of a *J. coenia* crispant displaying dual homeoses, *i.e.* with both FW➞HW (presumably due to *Ubx* gain-of-expression), and HW➞FW clones (akin to *Ubx* null mutations). Insets on the right describe forewing mutant clones (*IT1 mKO*), in apposition to CL-WT (contralateral forewings from the same individual), and WT HW (wild type hindwings from the same individual and mutant forewing side). **(D)** Examples of dual homeoses obtained when targeting orthologous sites in *V. cardui*.

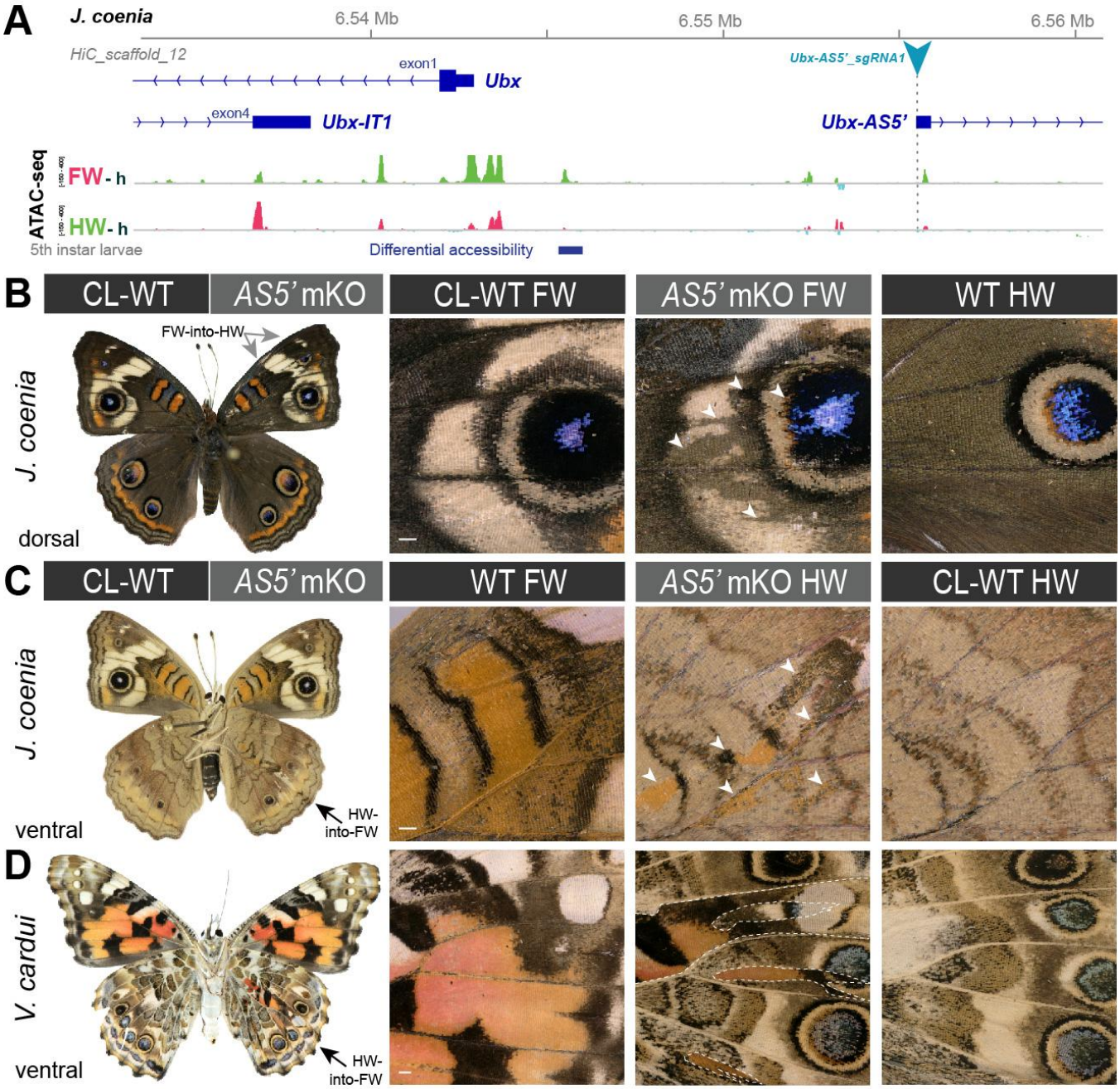


Figure 6. Homeoses obtained from CRISPR mutagenesis of the lncRNA *Ubx-AS5'* first exon. **(A)** CRISPR sgRNA targets (here shown in *J. coenia*), in the first exon of the non-coding *Ubx-AS5'* transcript. **(B)** A single *J. coenia* crispant showed a FW➞HW transformation. Insets on the right describe forewing mutant clones (*AS5' mKO*), in apposition to CL-WT (contralateral forewings from the same individual), and WT HW (wild-type hindwings from the same individual and mutant forewing side). **(C-D)** Examples of HW➞FW homeoses obtained in J. coenia or when targeting orthologous sites in *V. cardui*. Scale bars: 500 µm.

Dual effects of mutagenesis in a putative *Ubx cis*-regulatory module

In an attempt to probe for *Ubx* hindwing-specific regulatory sequences, we focused on a ~ 25kb region in the first intron of *Ubx* that displays an ATAC-seq signature of hindwing enrichment in open-chromatin relative to forewings, hereafter dubbed *CRM11* (**Fig. 7A**). We sub-divided this differentially accessible region into four peaks (*11a, b, c* and *d*). Targeting the ATAC-seq peaks with multiple sgRNAs spanning sub-domains *11a* and *11c* (*UbxCRE11a2c5* in *V. cardui*, *11a2a3c5c6* in *J. coenia*), or with a

290 single target in *11c* (*UbxCRE11c5* in *V. cardui*) yielded dual homeoses : FW→HW and HW→FW (**Figs. 7B-**
291 **D and S7**). Hindwing effects were reminiscent of *Ubx* protein coding knockouts (Tendolkar et al. 2021),
292 indicating that these crispant alleles with a hindwing phenotype produce *Ubx* loss-of-function effects.
293 Individuals with hindwing clones 2.75 times more common than individuals with forewings in this
294 dataset. Similarly to the lncRNA loci perturbation experiments, dual homeoses may indicate the
295 presence of hindwing activators and forewing repressors in the *CRM11* region, with various CRISPR
296 alleles producing a spectrum of indels and effects (see Discussion). It is noteworthy that while single-
297 target experiments showed little lethality (55% hatching rate), dual or quadruple injection mixes
298 resulted in low hatching rates of injected embryos (~ 10%). Multiple targeting thus appears to induce
299 high-rates of embryonic lethality, possibly due to chromosomal damage (Cullot et al. 2019; Zuccaro et
300 al. 2020). Dual targeting with *a2+c5* also yielded partial HW→FW homeoses in *V. cardui* under the form
301 of ectopic white eyespot foci phenotypes (**Fig. 7E**), as occasionally observed in *Ubx* null crispants
302 (Tendolkar et al. 2021), seemingly due to hypomorphic or heterozygous allelic states.
303 Next, we focused on a single target shared between both *V. cardui* and *J. coenia* in the *11b* sub-
304 domain. A whole genome alignment between 23 lepidopteran species and 2 trichopteran species
305 indicated that region *11b* is deeply conserved, suggesting important functional constraints on its
306 sequence (**Fig. S8A-B**). Mutagenesis of *11b* yielded a relatively high hatching rate (mean = 51.8 %),
307 indicating a rare occurrence of the deleterious mutational effects observed in multiple targeting (see
308 above). Four *J. coenia* crispants and two *V. cardui* crispants were obtained, all exclusively showing
309 hindwing phenotype devoid of forewing effects. HW→FW homeoses included a variety of phenotypes all
310 seen in *Ubx* CDS mutants (Tendolkar et al. 2021), including transformations of the orange Discalis
311 elements and the white band in *J. coenia*, and partial shifts in eyespot identity (**Fig. S8C**). Together the
312 consistency in direction of transformations and the deep conservation of the *11b* region suggests it may
313 encode an enhancer necessary for the transcriptional activation of *Ubx* in hindwings.

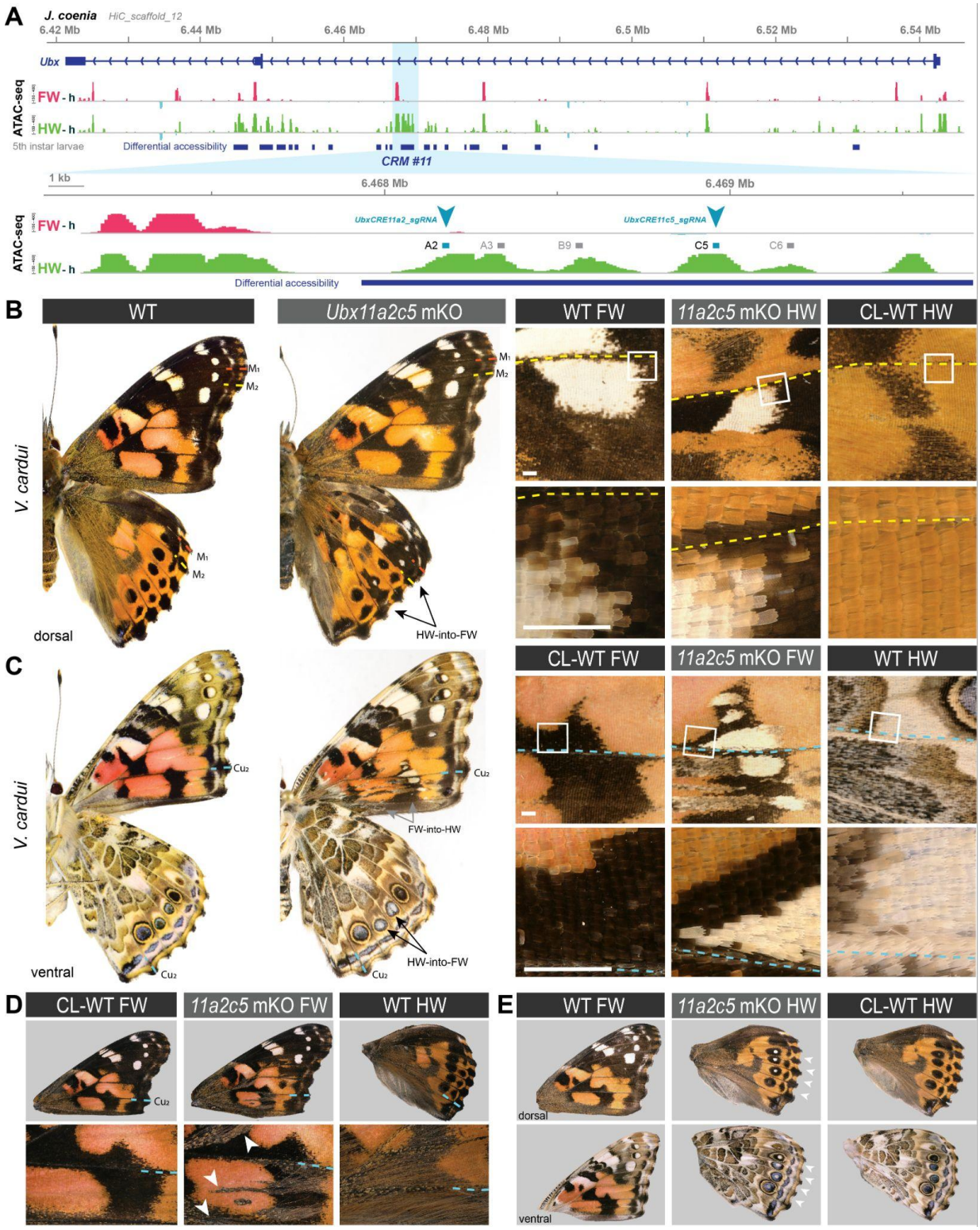


Figure 7. CRISPR perturbation of *Ubx CRM11* generates occasional dual homeotic phenotypes. (A) Overview of ATAC-seq differential chromatin accessibility profiles (hindwing - head tissues, green ; forewing - head tissue, magenta) across the *Ubx* first exon. Several regions show differential opening between wings, one of which (*CRM11*), was targeted here for CRISPR perturbation (sites *a2* and *c5* indicate sgRNA targets). **(B)** Dual homeosis phenotypes obtained in *V. cardui* following dual-targeting of *UbxCRE11a2c5*, including homeoses in color patterns and scale morphology. **(D)** Additional example of a *V. cardui UbxCRE11a2c5* crispant with a forewing phenotype (gain of hindwing hair patches, arrowheads). **(E)** Example of mild hindwing homeoses showing a white eyespot focus on the dorsal and ventral sides. These effects were

previously shown to occur in coding *Ubx* CRISPR knock-out experiments (Tendolkar et al, 2021). Contralateral (CL) WT wings are shown for comparison with mutant wings (B-E). Colored dashed lines: wing veins. Scale bars: 500 µm.

A sample of spontaneous homeotic mutants in *Heliconius* butterflies

Homeotic shifts between forewings and hindwings can occur naturally in Lepidoptera, and have been documented as pattern aberrations in museum specimens (Sibatani 1980, 1983). As a complement to CRISPR-induced homeoses, we document here a rich sample of forewing/hindwing homeotic mutants in the genus *Heliconius*, systematically collected by L. E. Gilbert between 1987 and 2016 in captive stocks at UT Austin, as well as in the wild. Across these 15 spontaneous mutants, 12 show HW→FW clones (**Fig. S9**), against 3 specimens with FW→HW effects (**Fig.8**). Mutant clones in this dataset were always posterior to the M_2 vein. Only 2 mosaic phenotypes were found on a dorsal side, with the 13 others appearing ventrally. These homeotic mosaics show pattern shifts with complete fore/hindwing conversions of scale types, as seen for instance in the loss of gray wing coupling scales on posterior ventral forewings (**Fig. 8A-B**), or conversely, in their acquisition in posterior hindwings (arrowheads in **Fig. S9B-D, H**). Homeoses also include noticeable local changes in wing shape, particularly in hindwings (asterisks in **Fig. S9**). Taken together, these effects are akin to CRISPR-induced perturbations at the *Ubx* locus. We speculate that fore/hindwing homeotic aberrations, found in nature and captive stocks, result from mutations at the *Ubx* locus itself.

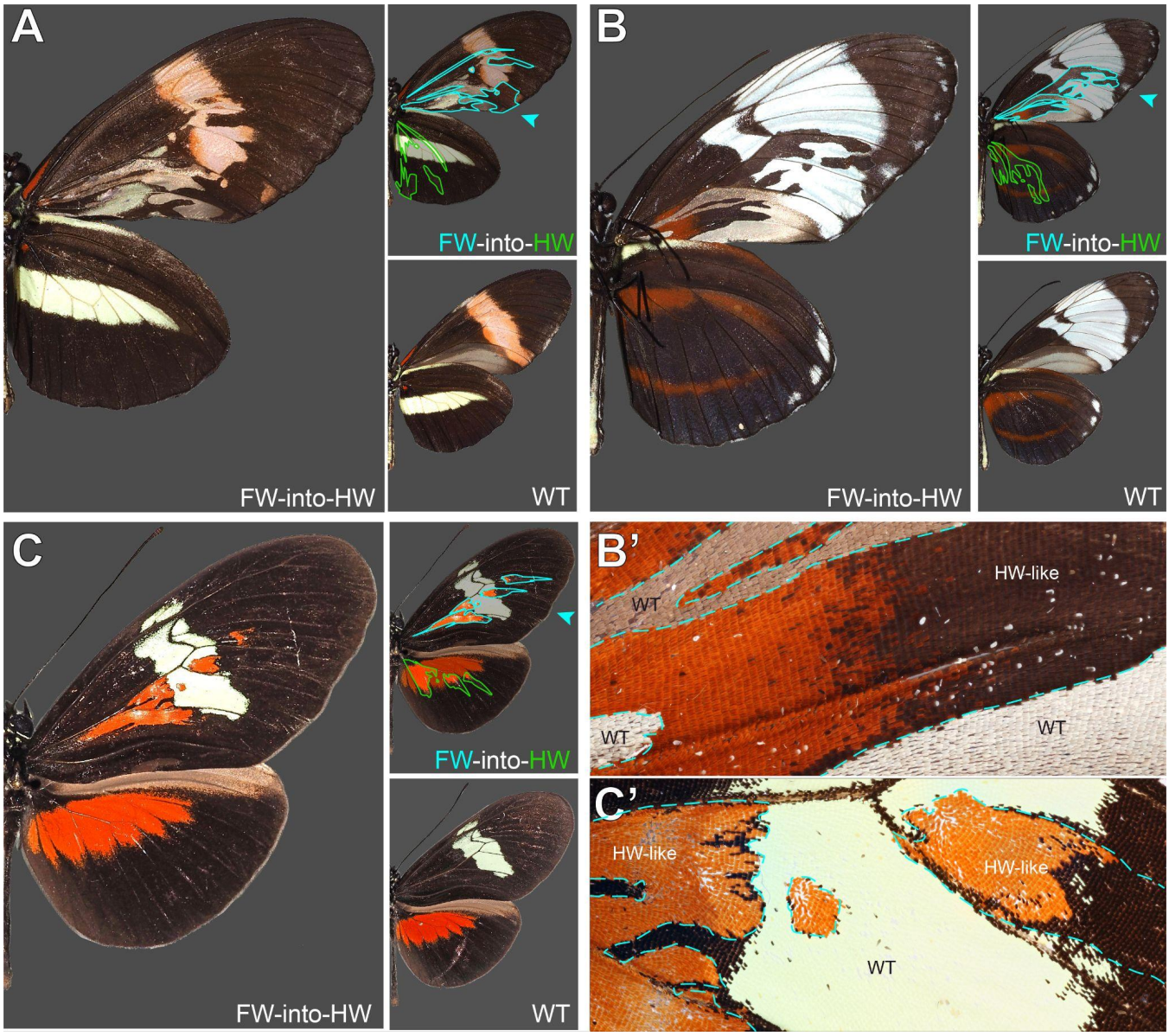


Figure 8. Mosaic forewing homeoses in *Heliconius* butterfly spontaneous mutants. Wild-type and mutant sides from the same individuals are shown in each panel, with one side digitally flipped to match left-to-right orientation. **A.** *Heliconius melpomene rosina*, ventral view. Wild-caught in the Osa Peninsula (Costa Rica), October 1989. **B.** *Heliconius cydno galanthus*, ventral view (magnified inset in **B'**). Stock culture from Organisation for Tropical Studies station, La Selva (Costa Rica), June 1990 **C.** *Heliconius himera*, dorsal view (magnified inset in **C'**). Stock Culture in the butterfly farm Heliconius Butterfly Works in Mindo (Ecuador), March 2008.

DISCUSSION

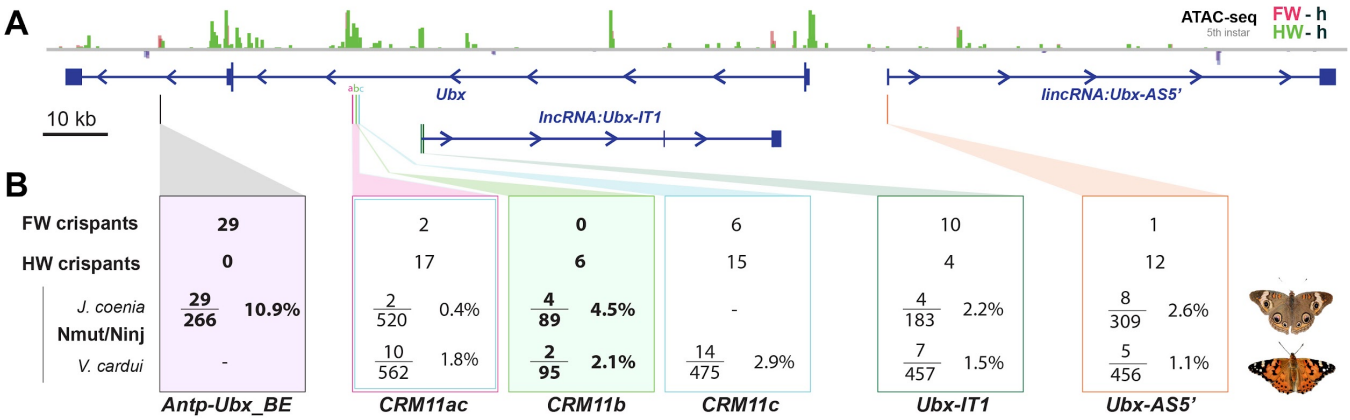


Figure 9. Summary of wing homeosis phenotypes obtained from mutational interrogation. (A) CRISPR targets at non-coding regions across the *Ubx* region, here visualized in *J. coenia*. **(B)** Summary of injection and adult phenotype data obtained across CRISPR experiments. FW/HW crispants : total number of individuals with forewing or hindwing homeotic clones, regardless of the injected species. Individuals with dual homeosis are counted in both categories. N_{mut}/N_{inj} : number of crispants obtained (N_{mut}), over the number of injected embryos for each species. Bold: experiments with consistent effects in only one segment. See **Table 1** for details.

An intronic region with ATAC-seq hindwing-enrichment regulates *Ubx*

All CRISPR targets yielded homeotic phenotypes (**Fig. 9**), with two kinds of interference with *Ubx* expression – forewing gain-of-function effects, and hindwing loss-of-function effects – and indicating the presence of regulatory sequences (broadly defined), that repress or enhance *Ubx* expression in this region. It is crucial here to highlight the limitations of the method, in order to derive proper insights about the functionality of the regulatory regions we tested. In essence, butterfly CRISPR experiments generate random mutations by non-homologous end joining repair, that are usually deletions (Connahs et al. 2019; Mazo-Vargas et al. 2022; Van Belleghem et al. 2023), and they require genotyping in a second (G_1) generation to be properly matched to a phenotype (genotyping G_0 mosaic wings is limited, because adult wings lost scale building cells that underlie a given phenotype). Previous data from other organisms suggests that Cas9 nuclease targeting can generate larger than expected mutations that evade common genotyping techniques (Shin et al. 2017; Adikusuma et al. 2018; Kosicki et al. 2018; Cullot et al. 2019; Owens et al. 2019). Even under the assumption that such mutations are relatively rare in butterfly embryos, the fact we injected >100 embryos in each experiment makes their occurrence likely (**Fig. 9**).

When targeting hindwing-enriched ATAC-seq peaks within the first intron of *Ubx* – from *CRM11* to the hindwing-enriched open-chromatin peak that coincides with the first exon of *Ubx-IT1* – we obtained a mixture of hindwing and forewing phenotypes. Given the potential heterogeneity of allele sizes in these experiments, it is difficult to conclude robustly about the function of individual targets. Nonetheless, the phenotypic data and in particular the obtention of dual homeoses suggest we disrupted sequences that are necessary to *Ubx* activation in hindwings, as well as to its repression in forewings. Bifunctional *cis*-regulatory elements that can switch between enhancer and silencer roles are prevalent in *Drosophila* (Gisselbrecht et al. 2020; Segert et al. 2021; Pang et al. 2022). The *CRM11* and *IT1* targets adjoin or overlap with open-chromatin signals in both wing sets (**Figs. 5A and 7A**), providing circumstantial evidence that these regions might serve as bifunctional elements. Similar observations were recently made in mutational interrogation experiments of the butterfly *WntA* patterning gene (Mazo-Vargas et al. 2022). Alternatively, silencers and enhancers may be tightly linked and interact in close proximity to shape gene expression (Méndez-González et al. 2023), implying in our

case that forewing and hindwing phenotypes are mediated by alleles spanning adjacent but distinct elements. A formal test of these mechanisms would require germline transmission and genotyping of these alleles, which was unsuccessful in our attempts at crossing *Ubx cis*-regulatory crispants.

In contrast with the dual effects obtained when targeting *CRM11a+c* (**Fig. 9**), *CRM11b* perturbation resulted in hindwing-limited effects, and may suggest that an *Ubx* enhancer was consistently compromised in this specific dataset. The high lethality and small size of mutant wing streaks suggest that only individuals with sparse, small mutant mitotic clones can survive to the adult stage. If this is true, *CRM11* may contain pleiotropic enhancers that are vital for normal *Ubx* function at earlier stages, but expression-reporter studies will be required to test this.

Parsing lncRNA-encoding regions – correlation or cause?

LncRNAs are emerging as important regulators of gene expression and developmental processes (Zhang et al. 2019; Statello et al. 2021). *IT1* targeting generated a majority of forewing phenotypes, suggesting perturbation of *Ubx* repression in the T2 segment. However, *IT1* showed low expression in forewing RNAseq datasets from *K. inachus,* and higher expression in the hindwing (**Fig. 1B**), a pattern inconsistent with a repressive role of the antisense *IT1* transcript on *Ubx* expression. It is generally challenging to disentangle the effects of transcription of a non-coding element from the potential effects of adjacent enhancers (Natoli and Andrau 2012; Pease et al. 2013). Therefore, an alternative explanation would be that the phenotypes are confounded by the overlap and proximity to open-chromatin regions, which may play *cis*-regulatory roles on *Ubx* via DNA-protein interactions, rather than via the lncRNA. If this is the case, it is possible that the targeted *Ubx-IT1* site, which yielded homeoses in both directions and bears both forewing and hindwing open-chromatin, is a bifunctional *cis*-regulatory element that can shift regulatory activities between these tissues (Gisselbrecht et al. 2020). Targeted mutagenesis of the *Ubx-AS5'* first exon mainly generated hindwing phenotypes, although with a relatively low-efficiency. Because this target is about 10 kb away from the *Ubx* promoter itself, it is plausible that the observed phenotypes were due to large deletions reaching the promoter region of *Ubx*. Because mutational interrogation alone cannot discern if phenotypic effects arose from regulatory failure at the chromatin or transcript level, determining if *AS5'* and *IT1* are functional lncRNAs will need further examination.

A TAD boundary element likely acts as an insulator preventing *Ubx* forewing expression

Tight maintenance of TAD boundaries at the *Hox* locus is crucial for accurate segment identity and is facilitated by insulator proteins (Stadler et al. 2017; Gambetta and Furlong 2018; Ramírez et al. 2018). The *Antp-Ubx_BE* element we targeted is in a good position to block interactions between *Antp* and *Ubx* (**Figs. 2-3**). Consistent with this idea, the last intron of *Ubx* contains 8 CTCF binding motifs that are all clustered within 5-kb around the forewing-enriched ATAC peak, including two sites at highly conserved positions that are only 100-bp away from the CRISPR target (**Fig. S1**). CTCF sites prevent cross-talk between regulatory domains in the fly BX-C complex, and result in *Hox* misexpression when deleted (Postika et al. 2018; Kyrchanova et al. 2020; Kaushal et al. 2022; Kahn et al. 2023). Thus, the density of CTCF sites in this region may be indicative of a *bona fide* insulator active in forewings.

CRISPR mutagenesis of *Antp-Ubx_BE* generated FW➞HW homeoses associated with a gain of UbdA antigenicity in forewings, with no effects in the other direction, in stark contrast with other targets (**Fig. 9B**). This suggests a possible de-insulation of the TAD boundary in the crispant clones, resulting in a TAD fusion or in a long-range interaction between a T2-specific enhancer and *Ubx* promoter. Similar

de-insulating effects of deletion alleles have been described at the *Notch* locus in *Drosophila* (Arzate-Mejía et al. 2020), in digit-patterning mutants in mice and humans (Lupiáñez et al. 2015; Anania et al. 2022), or at murine and fly *Hox* loci depleted of CTCF-mediated regulatory blocking (Narendra et al. 2015; Gambetta and Furlong 2018; Kyrchanova et al. 2020). It will be interesting to profile the binding of insulator proteins and transcriptional repressors across the butterfly *Hox* TAD landscape to shed more light onto the mechanisms of *Ubx* insulation, using *in vivo* assays (Bowman et al. 2014), or *in silico* predictions that take advantage of updated binding matrices for insect insulator proteins (Mitra et al. 2018). Of note, our CRISPR target is adjacent to an hindwing-enriched peak that also presented CTCF binding sites (**Fig. 4A**). Following a similar logic, this site could be a candidate insulator specific to *Ubx*-expressing tissues like the hindwing, a hypothesis that will require further testing.

Making sense of spontaneous wing homeotic mutants

In this article, we documented a large sample of spontaneous homeotic mutants obtained in *Heliconius spp.* All homeotic clones were limited to the wing posterior compartments (*ie.* posterior to the M_2 vein), possibly because of parasegmental, compartment-specific regulatory domains that played historic roles in the study of *Drosophila* BX-C regulation (Maeda and Karch 2015). Sibatani documented in Lepidoptera that "*the mosaics of F/H homeosis tend to occur most frequently in the posterior half of the wing, the boundary of the anterior and posterior halves occurring somewhere in space* M_1-M_2" (Sibatani 1983). Our collection of spontaneous *Heliconius* mutants only displayed clones in posterior regions, consistently with this trend. However, our CRISPR perturbation assays of *J. coenia* and *V. cardui* *cis*-regulatory regions also generated anterior clones, with all targets. Deciphering how butterfly *Ubx* regulation is compartmentized between parasegmental or wing antero-posterior domains will require additional investigation. Most *Heliconius* homeoses were in the hindwings (*ie.* putative *Ubx* loss-of-expression clones), and among these, all but one were ventral (**Fig. S9**). Three mutants showed forewing homeoses (*ie.* putative *Ubx* gain-of-expression clones), two of them ventral and one of them dorsal (**Fig. 8**). The systematic reviews of wing homeosis in Lepidoptera found that forewing homeoses are almost as common as hindwing ones (Sibatani 1980, 1983). Our mutational interrogation assays, while coarse in nature, revealed the existence of activating and repressing *cis*-regulatory sequences at the *Ubx* locus itself. Spontaneous FW↔HW homeoses observed in butterflies and moths may thus result from somatic mutations or transposition events at this locus.

MATERIALS AND METHODS

Genome annotations and transcriptomic analysis

Nymphalid genome sequences of the *Hox* cluster and their annotations were extracted from the NCBI Assembly and Lepbase online repositories (Challis et al. 2016; Kitts et al. 2016) as follows : *V. cardui* from NCBI *ilVanCard2.1* and LepBase *Vc_v1* ; *A.* (*Nymphalis*) *io* from NCBI *ilAglIoxx1.1*; *J. coenia* from Lepbase *Jc_v2*; *P xylostella* from NCBI *ilPluXylo3.1*. The *Ubx* regions from *ilVanCard2.2, Vc_v1, and Jc_v2* were manually re-annotated using wing transcriptome data on the NCBI SRA (BioProjects *PRJNA661999*, *PRJNA293289*, *PRJNA237755*, *PRJNA385867*, and *PRJNA498283*) The genome sequence of *K. inachus* was obtained from the Dryad repository (Yang et al. 2020). Differential gene expression analysis across the *K. inachus Ubx* locus were carried out using wing transcriptome data available on the NCBI SRA (BioProject *PRJNA698433*), following a manual re-annotation of a

469 preliminary gene models provided by Peiwen Yang and Wei Zhang (Wang et al. 2022). All transcripts
470 analyses were performed using the *STAR* intron-aware aligner and *DEseq2* expression analysis package
471 as previously described (Love et al. 2014; Dobin and Gingeras 2016; Hanly et al. 2019, 2022). Expression
472 levels were calculated as genome-wide normalized counts and pairwise Wald tests were performed to
473 assess differential expression between forewings and hindwings. Multiple test adjustment was
474 performed using Benjamini and Hochberg correction.
475
476 **Hi-C and ATAC-seq analyses**
477 Hi-C data from *J. coenia* fifth instar larval forewings and 48-72 h APF pupa hindwings are available at
478 the NCBI SRA BioProject *PRJNA641138* (van der Burg et al. 2020). Triplicated ATAC-seq datasets for
479 larval and pupal wing and head tissues of *J. coenia* and *V. cardui* (van der Burg et al. 2019, 2020; Mazo-
480 Vargas et al. 2022) are available on the NCBI SRA BioProjects *PRJNA497878*, *PRJNA695303*, and
481 *PRJNA559165*. All the ATAC-seq and Hi-C data were re-analysed on *J. coenia* and *V. cardui Ubx*
482 genomic regions as previously described (Mazo-Vargas et al. 2022). Briefly, matrices of interactions
483 were constructed by mapping paired reads against the *Junonia coenia* genome (Mazo-Vargas et al.,
484 2022) using *hicBuildMatrix* (Ramírez et al. 2018). Matrices from larvae and pupae were normalized
485 using *hicNormalize* and corrected with the Knight-Ruiz matrix balancing algorithm.The definitions of
486 topologically associating domains (TADs) can be influenced by various factors such as the choice of
487 software, parameters, sequencing depth, and the presence of experimental noise. To ensure reliability, it
488 is recommended to compare TAD calls with independent datasets, such as histone marks or known
489 factors associated with TAD boundaries. In the absence of these specific datasets, we employed a
490 different combination of parameters in the *hicFindTADs* tool and compared the resulting TAD calls. HiC
491 matrices at 10 kb and 20 kb bin resolutions were utilized, and TAD insulation scores were evaluated
492 using a false-discovery rate correction for multiple testing, with *p-value* thresholds of 0.01 and 0.005.
493 Consistent TAD boundary calls with negative TAD separation scores were selected to define domain
494 limits at 10 kb and 20 kb matrix resolutions.
495
496 **CTCF motif binding predictions**
497 The program *fimo* was used to scan for the *J. coenia* candidate TAD boundary region
498 (HiC_scaffold_12:6430000-6444000) for canonical CTCF binding sites, using the positional weight
499 matrix MA0205.1 deposited in the JASPAR database (Holohan et al. 2007; Cuellar-Partida et al. 2012;
500 Castro-Mondragon et al. 2022).
501
502 **Genomic conservation analyses**
503 We generated whole-genome alignments of 25 Lepidoptera and 2 Trichoptera reference species from
504 NCBI Assembly using *ProgressiveCactus* (Armstrong et al. 2020), and *HAL tools* (Hickey et al. 2013) for
505 converting the resulting HAL file to the MAF format. We provided a species topology tree of 23
506 Lepidoptera species to *PhyloFit* (Hubisz et al. 2011) to fit a multiple sequence alignment on the
507 reference *J. coenia Ubx* locus, using *HKY85* as the substitution model. Using *PhastCons* (Siepel et al.
508 2005), we then generated conservation score plots using standard parameters (target-coverage = 0.45;
509 expected-length = 12; rho = 0.1) stored in BED and WIG file formats.
510
511 **Butterfly rearing and CRISPR microinjections**
512 *J. coenia* and *V. cardui* colonies were maintained at 25°C and 60-70% relative humidity in a growth
513 chamber with a 14:10 light:dark photoperiod. Larval rearing on artificial diets, egg collection, and

514 microinjections followed previously described methods (Martin et al. 2020; Tendolkar et al. 2021).
515 Cas9:sgRNA heteroduplexes were prepared as previously described (Martin et al., 2020). Frozen
516 aliquots of Cas9-2xNLS (2.5 µL ; 1,000 ng/µL) and sgRNA (2.5 µL ; 500 ng/µL) were mixed in 2:1 and
517 4:1:1 mass ratios for single and dual target injections, respectively. CRISPR sgRNA targets are listed in
518 **Table S1**.
519
520 **Antibody stainings**
521 Fifth instar wing disks were dissected in ice cold Phosphate Buffer Saline (PBS), fixed for 15-20 min at
522 room temperature in methanol-free formaldehyde diluted to 4% in PBS / 2mM EGTA (egtazic acid),
523 washed in PBS with 0.1% Triton X-100 (PT), stored in PT with 0.5% Bovine Serum albumin (PT-BSA),
524 incubated overnight at 4°C in PT-BSA with a 1:5 dilution of the FP6.87 antibody serum (mouse
525 monoclonal, Developmental Studies Hybridoma Bank), and washed in PT. A 1:250 dilution of anti-Mouse
526 IgG antibody coupled to AlexaFluor488 or Rabbit AlexaFluor555 was made in PT-BSA and spun down
527 at 14,000 rcf to remove aggregates, and incubated with wings for 2 h at room temperature, before
528 additional washes, incubation in 50% glycerol-PBS with DAPI (4′,6-diamidino-2-phenylindole) nuclear
529 stain, and incubation and mounting in 60% glycerol-PBS with 2mM of EDTA
530 (Ethylenediaminetetraacetic acid).
531
532 **Imaging**
533 Full-mount photographs of *J. coenia* and *V. cardui* were taken on a Nikon D5300 digital camera mounted
534 with an AF-S VR MicroNikkor 105mm f/2.8G lens, with magnified views taken using a Keyence VHX-5000
535 digital microscope mounted with VH-Z00T and VH-Z100T lenses. Immunofluorescent stainings were
536 imaged on an Olympus BX53 epifluorescent microscope mounted with UPLFLN 4x/0.13 and 10X/0.3
537 objectives.
538
539 **ACKNOWLEDGEMENTS**
540 We thank Ling Sheng Loh and the undergraduate researchers from the Martin Lab for assistance with
541 micro-injections and insect rearing, Rachel Canalichio and the GWU Harlan Greenhouse personnel for
542 growing host plants, Patricia Hernandez for sharing microscopes, and Alex Wild for assistance with
543 *Heliconius* microphotographs at UT Austin. We wish to acknowledge James Lewis and Bob Reed for
544 stimulating insights on open-chromatin biology and the *Hox* locus, as well as for generating Hi-C
545 libraries published in previous publications that we re-analyzed here. This work was supported by the
546 NSF awards IOS-1656553 and IOS-2110534 to AM, the Wilbur V. Harlan Research Fellowship to AT, the
547 NSF Postdoctoral Research Fellowship in Biology to AMV, and the Smithsonian Institution Biodiversity
548 Genomics Fellowship to JJH.

549 **REFERENCES**

550 Adikusuma F, Piltz S, Corbett MA, et al (2018) Large deletions induced by Cas9 cleavage. Nature 560:E8–E9
551 Anania C, Acemel RD, Jedamzick J, et al (2022) In vivo dissection of a clustered-CTCF domain boundary reveals
552 developmental principles of regulatory insulation. Nature Genetics 54:1026–1036
553 Armstrong J, Hickey G, Diekhans M, et al (2020) Progressive Cactus is a multiple-genome aligner for the
554 thousand-genome era. Nature 587:246–251. https://doi.org/10.1038/s41586-020-2871-y
555 Arzate-Mejía RG, Josué Cerecedo-Castillo A, Guerrero G, et al (2020) In situ dissection of domain boundaries
556 affect genome topology and gene transcription in Drosophila. Nature communications 11:894
557 Bowman SK, Deaton AM, Domingues H, et al (2014) H3K27 modifications define segmental regulatory domains
558 in the Drosophila bithorax complex. Elife 3:e02833
559 Buffry AD, Kittelmann S, McGregor AP (2023) Characterisation of the role and regulation of Ultrabithorax in
560 sculpting fine-scale leg morphology. Frontiers in Cell and Developmental Biology 11:
561 Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, et al (2022) JASPAR 2022: the 9th release of the
562 open-access database of transcription factor binding profiles. Nucleic acids research 50:D165–D173
563 Challis RJ, Kumar S, Dasmahapatra KKK, et al (2016) Lepbase: the Lepidopteran genome database. bioRxiv
564 056994
565 Connahs H, Tlili S, van Creij J, et al (2019) Activation of butterfly eyespots by Distal-less is consistent with a
566 reaction-diffusion process. Development 146:dev169367
567 Cuellar-Partida G, Buske FA, McLeay RC, et al (2012) Epigenetic priors for identifying active transcription factor
568 binding sites. Bioinformatics 28:56–62
569 Cullot G, Boutin J, Toutain J, et al (2019) CRISPR-Cas9 genome editing induces megabase-scale chromosomal
570 truncations. Nature communications 10:1136
571 Dobin A, Gingeras TR (2016) Optimizing RNA-Seq mapping with STAR. In: Data mining techniques for the life
572 sciences. Springer, pp 245–262
573 Feng S, Rastogi C, Loker R, et al (2022) Transcription factor paralogs orchestrate alternative gene regulatory
574 networks by context-dependent cooperation with multiple cofactors. Nature Communications 13:1–19
575 Ferguson L, Marletaz F, Carter J-M, et al (2014) Ancient expansion of the Hox cluster in Lepidoptera generated
576 four homeobox genes implicated in extra-embryonic tissue formation. PLoS Genet 10:e1004698
577 Gambetta MC, Furlong EE (2018) The insulator protein CTCF is required for correct Hox gene expression, but not
578 for embryonic development in Drosophila. Genetics 210:129–136
579 Garaulet DL, Lai EC (2015) Hox miRNA regulation within the Drosophila Bithorax complex: Patterning behavior.
580 Mechanisms of development 138:151–159
581 Gaunt SJ (2022) Seeking sense in the Hox gene cluster. Journal of Developmental Biology 10:48
582 Gisselbrecht SS, Palagi A, Kurland JV, et al (2020) Transcriptional silencers in drosophila serve a dual role as
583 transcriptional enhancers in alternate cellular contexts. Molecular cell 77:324–337
584 Guo Y, Xu Q, Canzio D, et al (2015) CRISPR inversion of CTCF sites alters genome topology and
585 enhancer/promoter function. Cell 162:900–910
586 Gutierrez-Perez I, Rowley MJ, Lyu X, et al (2019) Ecdysone-Induced 3D chromatin reorganization involves active
587 enhancers bound by pipsqueak and polycomb. Cell reports 28:2715–2727
588 Hajirnis N, Mishra RK (2021) Homeotic Genes: Clustering, Modularity, and Diversity. Frontiers in Cell and
589 Developmental Biology 9:
590 Hanly JJ, Livraghi L, Heryanto C, et al (2022) A large deletion at the cortex locus eliminates butterfly wing
591 patterning. G3 12:jkac021
592 Hanly JJ, Wallbank RW, McMillan WO, Jiggins CD (2019) Conservation and flexibility in the gene regulatory
593 landscape of heliconiine butterfly wings. Evodevo 10:1–14
594 Heffer A, Pick L (2013) Conservation and variation in Hox genes: how insect models pioneered the evo-devo field.
595 Annu Rev Entomol 58:161–179
596 Hermann A, Kosman D, McGinnis W, Tour E (2022) The expression of Drosophila melanogaster Hox gene
597 Ultrabithorax is not overtly regulated by the intronic long noncoding RNA lncRNA:PS4 in a wild-type
598 genetic background. G3 Genes|Genomes|Genetics 12:jkab374. https://doi.org/10.1093/g3journal/jkab374
599 Hickey G, Paten B, Earl D, et al (2013) HAL: a hierarchical format for storing and analyzing multiple genome
600 alignments. Bioinformatics 29:1341–1342. https://doi.org/10.1093/bioinformatics/btt128
601 Holohan EE, Kwong C, Adryan B, et al (2007) CTCF genomic binding sites in Drosophila and the organisation of
602 the bithorax complex. PLoS genetics 3:e112
603 Hubisz MJ, Pollard KS, Siepel A (2011) PHAST and RPHAST: phylogenetic analysis with space/time models.
604 Brief Bioinform 12:41–51. https://doi.org/10.1093/bib/bbq072
605 Hughes CL, Kaufman TC (2002) Hox genes and the evolution of the arthropod body plan. Evolution &
606 development 4:459–499

607 Ibragimov A, Bing XY, Shidlovskii Y, et al (2022) The insulating activity of the Drosophila BX-C chromatin
608 boundary Fub-1 is parasegmentally regulated by lncRNA read-through. bioRxiv
609 Kahn TG, Savitsky M, Kuong C, et al (2023) Topological screen identifies hundreds of Cp190-and CTCF-
610 dependent Drosophila chromatin insulator elements. Science Advances 9:eade0090
611 Kaushal A, Dorier J, Wang B, et al (2022) Essential role of Cp190 in physical and regulatory boundary formation.
612 Science Advances 8:eabl8834
613 Kawahara AY, Plotkin D, Espeland M, et al (2019) Phylogenomics reveals the evolutionary timing and pattern of
614 butterflies and moths. Proceedings of the National Academy of Sciences 116:22657–22663
615 Kitts PA, Church DM, Thibaud-Nissen F, et al (2016) Assembly: a resource for assembled genomes at NCBI.
616 Nucleic acids research 44:D73–D80
617 Kosicki M, Tomberg K, Bradley A (2018) Repair of double-strand breaks induced by CRISPR–Cas9 leads to large
618 deletions and complex rearrangements. Nature biotechnology 36:765–771
619 Kyrchanova O, Maksimenko O, Ibragimov A, et al (2020) The insulator functions of the Drosophila polydactyl
620 C2H2 zinc finger protein CTCF: Necessity versus sufficiency. Science advances 6:eaaz3152
621 Lewis DL, DeCamillis MA, Brunetti CR, et al (1999) Ectopic gene expression and homeotic transformations in
622 arthropods using recombinant Sindbis viruses. Current Biology 9:1279–1287.
623 https://doi.org/10.1016/S0960-9822(00)80049-4
624 Lewis JJ, Reed RD (2018) Genome-wide regulatory adaptation shapes population-level genomic landscapes in
625 Heliconius. Molecular biology and evolution
626 Livraghi L (2017) Hox3 duplication and divergence in the Lepidoptera. PhD Thesis, Oxford Brookes University
627 Livraghi L, Martin A, Gibbs M, et al (2017) CRISPR/Cas9 as the Key to Unlocking the Secrets of Butterfly Wing
628 Pattern Development and Its Evolution. Advances in Insect Physiology
629 Lohse K, Mackintosh A, Vila R, Consortium DT of L (2021a) The genome sequence of the European peacock
630 butterfly, Aglais io (Linnaeus, 1758). Wellcome Open Research 6:258
631 Lohse K, Wright C, Talavera G, et al (2021b) The genome sequence of the painted lady, Vanessa cardui
632 Linnaeus 1758. Wellcome Open Research 6:324
633 Loker R, Sanner JE, Mann RS (2021) Cell-type-specific Hox regulatory strategies orchestrate tissue identity.
634 Current Biology 31:4246–4255
635 Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with
636 DESeq2. Genome biology 15:1–21
637 Lupiáñez DG, Kraft K, Heinrich V, et al (2015) Disruptions of topological chromatin domains cause pathogenic
638 rewiring of gene-enhancer interactions. Cell 161:1012–1025
639 Maeda RK, Karch F (2015) The open for business model of the bithorax complex in Drosophila. Chromosoma
640 124:293–307
641 Mahfooz N, Turchyn N, Mihajlovic M, et al (2007) Ubx regulates differential enlargement and diversification of
642 insect hind legs. PLoS One 2:e866
643 Mallo M, Alonso CR (2013) The regulation of Hox gene expression during animal development. Development
644 140:3951–3963
645 Martin A, Wolcott NS, O'Connell LA (2020) Bringing immersive science to undergraduate laboratory courses
646 using CRISPR gene knockouts in frogs and butterflies. Journal of Experimental Biology 223:
647 Masumoto M, Yaginuma T, Niimi T (2009) Functional analysis of Ultrabithorax in the silkworm, Bombyx mori,
648 using RNAi. Development genes and evolution 219:437–444
649 Matsuoka Y, Monteiro A (2018) Melanin pathway genes regulate color and morphology of butterfly wing scales.
650 Cell reports 24:56–65
651 Matsuoka Y, Monteiro A (2021) Hox genes are essential for the development of eyespots in Bicyclus anynana
652 butterflies. Genetics 217:iyaa005
653 Matsuoka Y, Monteiro A (2022) Ultrabithorax modifies a regulatory network of genes essential for butterfly
654 eyespot development in a wing sector-specific manner. Development 149:dev200781
655 Matsuoka Y, Murugesan SN, Prakash A, Monteiro A (2022) Lepidopteran prolegs are novel traits, not leg
656 homologs. bioRxiv
657 Mazo-Vargas A, Langmüller AM, Wilder A, et al (2022) Deep cis-regulatory homology of the butterfly wing pattern
658 ground plan. Science 378:304–308
659 Méndez-González ID, Williams TM, Rebeiz M (2023) Changes in locus wide repression underlie the evolution of
660 Drosophila abdominal pigmentation. PLoS genetics 19:e1010722
661 Merabet S, Carnesecchi J (2022) Hox dosage and morphological diversification during development and
662 evolution. In: Seminars in Cell & Developmental Biology. Elsevier
663 Mitra S, Biswas A, Narlikar L (2018) DIVERSITY in binding, regulation, and evolution revealed from high-
664 throughput ChIP. PLoS Computational Biology 14:e1006090
665 Moniot-Perron L, Moindrot B, Manceau L, et al (2023) The Drosophila Fab-7 boundary modulates Abd-B gene

activity by guiding an inversion of collinear chromatin organization and alternate promoter use. Cell Reports 42:111967
Mulhair P, Crowley L, Boyes DH, et al (2022) Diversity, duplication, and genomic organization of homeobox genes in Lepidoptera. Genome Research gr-277118
Mulhair PO, Holland PW (2022) Evolution of the insect Hox gene cluster: Comparative analysis across 243 species. In: Seminars in Cell & Developmental Biology. Elsevier
Narendra V, Rocha PP, An D, et al (2015) CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science 347:1017–1021
Natoli G, Andrau J-C (2012) Noncoding transcription at enhancers: general principles and functional models. Annual review of genetics 46:1–19
Nijhout HF, Rountree DB (1995) Pattern induction across a homeotic boundary in the wings of Precis coenia (Hbn.) (Lepidoptera: Nymphalidae). International Journal of Insect Morphology and Embryology 24:243–251
Owens DD, Caulder A, Frontera V, et al (2019) Microhomologies are prevalent at Cas9-induced larger deletions. Nucleic acids research 47:7402–7417
Pace RM, Grbić M, Nagy LM (2016) Composition and genomic organization of arthropod Hox clusters. EvoDevo 7:1–11
Pang B, van Weerd JH, Hamoen FL, Snyder MP (2022) Identification of non-coding silencer elements and their regulation of gene expression. Nature Reviews Molecular Cell Biology 1–13
Paul R, Giraud G, Domsch K, et al (2021) Hox dosage contributes to flight appendage morphology in Drosophila. Nature Communications 12:2892
Pease B, Borges AC, Bender W (2013) Noncoding RNAs of the Ultrabithorax domain of the Drosophila bithorax complex. Genetics 195:1253–1264
Postika N, Metzler M, Affolter M, et al (2018) Boundaries mediate long-distance interactions between enhancers and promoters in the Drosophila Bithorax complex. PLoS genetics 14:e1007702
Prasad N, Tarikere S, Khanale D, et al (2016) A comparative genomic analysis of targets of Hox protein Ultrabithorax amongst distant insect species. Scientific reports 6:27885
Ramírez F, Bhardwaj V, Arrigoni L, et al (2018) High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nature communications 9:189
Refki PN, Armisén D, Crumière AJJ, et al (2014) Emergence of tissue sensitivity to Hox protein levels underlies the evolution of an adaptive morphological trait. Developmental biology 392:441–453
Savitsky M, Kim M, Kravchuk O, Schwartz YB (2016) Distinct roles of chromatin insulator proteins in control of the Drosophila bithorax complex. Genetics 202:601–617
Seal RL, Tweedie S, Bruford EA (2022) A standardised nomenclature for long non-coding RNAs. IUBMB life
Segert JA, Gisselbrecht SS, Bulyk ML (2021) Transcriptional silencers: Driving gene expression with the brakes on. Trends in Genetics 37:514–527
Shin HY, Wang C, Lee HK, et al (2017) CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. Nature communications 8:15464
Shippy TD, Ronshaugen M, Cande J, et al (2008) Analysis of the Tribolium homeotic complex: insights into mechanisms constraining insect Hox clusters. Development genes and evolution 218:127–139
Sibatani A (1980) Wing homoeosis in Lepidoptera: a survey. Developmental Biology 79:1–18
Sibatani A (1983) A compilation of data on wing homoeosis in Lepidoptera. J Res Lepidop 22:1–46
Siepel A, Bejerano G, Pedersen JS, et al (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 15:1034–1050. https://doi.org/10.1101/gr.3715005
Srinivasan A, Mishra RK (2020) Lessons on gene regulation learnt from the Drosophila melanogaster bithorax complex. International Journal of Developmental Biology 64:151–158
Stadler MR, Haines JE, Eisen MB (2017) Convergence of topological domain boundaries, insulators, and polytene interbands revealed by high-resolution mapping of chromatin contacts in the early Drosophila melanogaster embryo. Elife 6:e29550
Statello L, Guo C-J, Chen L-L, Huarte M (2021) Gene regulation by long non-coding RNAs and its biological functions. Nature reviews Molecular cell biology 22:96–118
Tendolkar A, Pomerantz AF, Heryanto C, et al (2021) Ultrabithorax Is a Micromanager of Hindwing Identity in Butterflies and Moths. Front Ecol Evol 9:. https://doi.org/10.3389/fevo.2021.643661
Thomas JA, Frandsen PB, Prendini E, et al (2020) A multigene phylogeny and timeline for Trichoptera (Insecta). Systematic Entomology 45:670–686
Tomoyasu Y (2017) Ultrabithorax and the evolution of insect forewing/hindwing differentiation. Current opinion in insect science 19:8–15
Tong X, Hrycaj S, Podlaha O, et al (2014) Over-expression of Ultrabithorax alters embryonic body plan and wing patterns in the butterfly Bicyclus anynana. Developmental biology 394:357–366

Tong XL, Fu MY, Chen P, et al (2017) Ultrabithorax and abdominal-A specify the abdominal appendage in a dosage-dependent manner in silkworm, Bombyx mori. Heredity 118:578

Van Belleghem SM, Ruggieri AA, Concha C, et al (2023) High level of novelty under the hood of convergent evolution. Science 379:1043–1049

van der Burg KR, Lewis JJ, Brack BJ, et al (2020) Genomic architecture of a genetically assimilated seasonal color pattern. Science 370:721–725

van der Burg KR, Lewis JJ, Martin A, et al (2019) Contrasting roles of transcription factors Spineless and EcR in the highly dynamic chromatin landscape of butterfly wing metamorphosis. Cell reports 27:1027–1038

Wang H, Hu H, Xiang Z, et al (2019) Identification and characterization of a new long noncoding RNA iab-1 in the Hox cluster of silkworm, Bombyx mori identification of iab-1. Journal of Cellular Biochemistry 120:17283–17292. https://doi.org/10.1002/jcb.28990

Wang H, Tong X, Liu M, et al (2017) Fine mapping of a degenerated abdominal legs mutant (Edl) in silkworm, Bombyx mori. Plos one 12:e0169224

Wang S, Teng D, Li X, et al (2022) The evolution and diversification of oakleaf butterflies. Cell 185:3138–3152

Warren RW, Nagy L, Selegue J, et al (1994) Evolution of homeotic gene regulation and function in flies and butterflies. Nature 372:458

Weatherbee SD, Frederik Nijhout H, Grunert LW, et al (1999) Ultrabithorax function in butterfly wings and the evolution of insect wing patterns. Current Biology 9:109–115. https://doi.org/10.1016/S0960-9822(99)80064-5

Weatherbee SD, Halder G, Kim J, et al (1998) Ultrabithorax regulates genes at several levels of the wing-patterning hierarchy to shape the development of the Drosophila haltere. Genes & Development 12:1474–1482. https://doi.org/10.1101/gad.12.10.1474

Wootton RJ (1993) Leading edge section and asymmetric twisting in the wings of flying butterflies (Insecta, Papilionoidea)

Yang J, Wan W, Xie M, et al (2020) Chromosome-level reference genome assembly and gene editing of the dead-leaf butterfly Kallima inachus. Molecular Ecology Resources 20:1080–1092

Zhang L, Steward RA, Wheat CW, Reed RD (2021) High-quality genome assembly and comprehensive transcriptome of the painted lady butterfly Vanessa cardui. Genome biology and evolution 13:evab145

Zhang X, Wang W, Zhu W, et al (2019) Mechanisms and functions of long non-coding RNAs at multiple regulatory levels. International journal of molecular sciences 20:5573

Zheng Z, Khoo A, Fambrough Jr D, et al (1999) Homeotic gene expression in the wild-type and a homeotic mutant of the moth Manduca sexta. Development genes and evolution 209:460–472

Zuccaro MV, Xu J, Mitchell C, et al (2020) Allele-specific chromosome removal after Cas9 cleavage in human embryos. Cell 183:1650–1664

SUPPLEMENTARY INFORMATION

Species	sgRNA name	Target Sequence (5' to 3') PAM sequence not shown
J. coenia	*Antp-Ubx_BE*	CTCGAATATGGAGATATCGG
	UbxCRE11a3	ACGGACCTCCGCTTTCCTGG
	UbxCRE11c6	AACTGGTGCAGTGCCTTGTA
J. coenia *+ V. cardui*	*UbxCRE11a2*	CTACTCTGTTCGGACATTCG
	UbxCRE11c5	GCTGCCGCGAGTCTGAATCG
	UbxCRE11b9	TTCATGTATGAACCATGACG
	UbxIT1#1	CCTTCGCATAAGTTCGGATAGG
	Bxd1	TATCGGTCGTTCGTCACACA
V. cardui	*UbxIT1#2*	CTCGGCTATGTGTCGAGGGC

Table S1. List of sgRNAs used in CRISPR experiments.

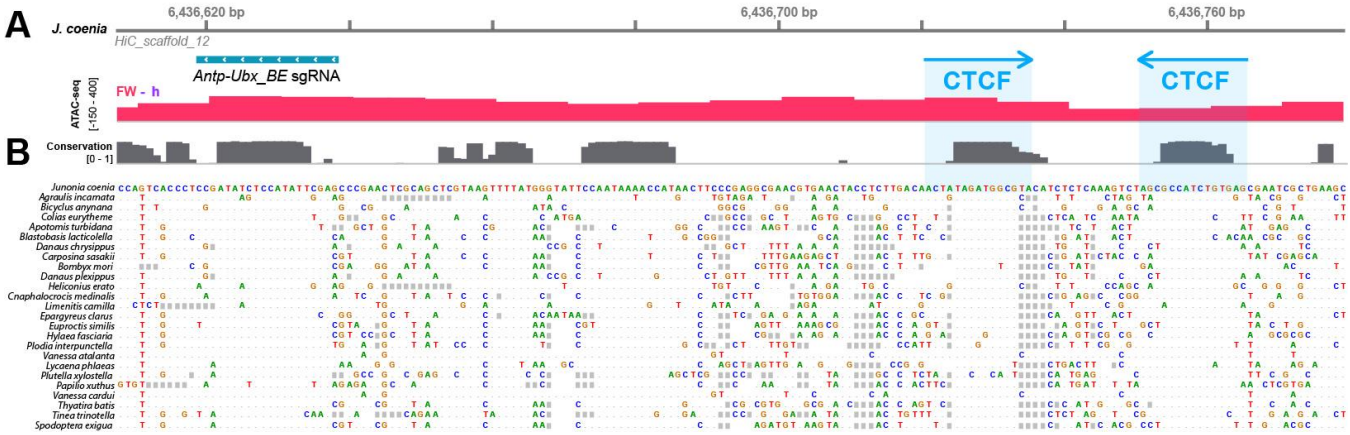


Figure S1. Prediction of two conserved CTCF binding sites at *Antp-Ubx_BE*. (**A**) Sequence-level view of a 180-bp genomic interval including the *Antp-Ubx_BE* sgRNA (turquoise) in *J. coenia*, overlapping with an ATAC-seq peak of forewing-enriched chromatin opening (red). The CRISPR target is about 100 bp away from two predicted binding sites for the *Drosophila* CTCF insulator protein. (**B**) High-level of nucleotide conservation at thesgRNA site and CTCF motifs across Lepidoptera and Trichoptera representative genomes, indicative of functional constraints on these sequences.

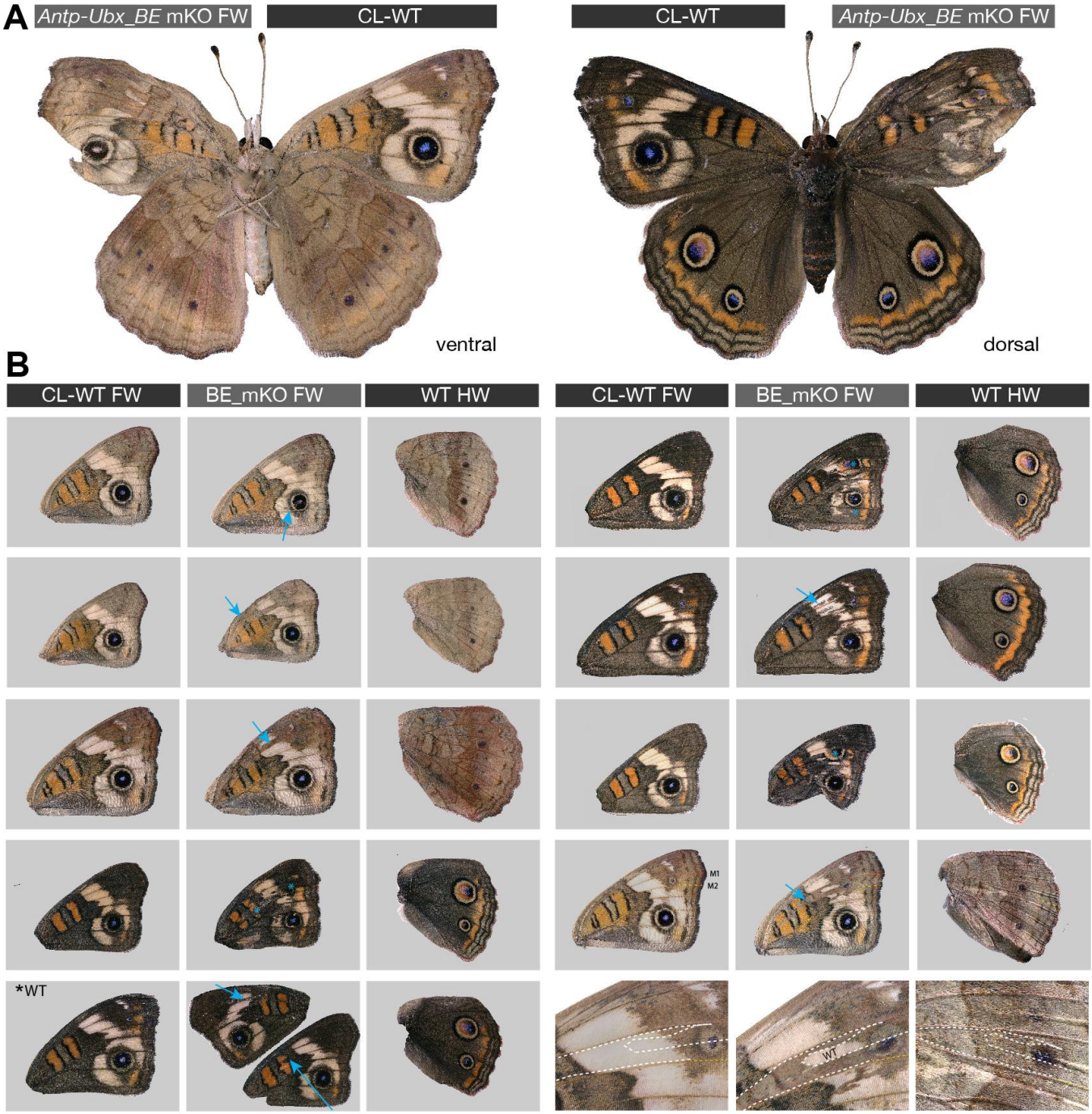


Figure S2. CRISPR perturbation of the *Antp-Ubx* boundary element results in FW-to-HW homeosis. (**A**) Example of an *Antp-Ubx_BE* crispant with a unilateral phenotype on the right forewing. (**B**) Additional examples of forewing homeoses in *Antp-Ubx_BE* crispant. Wing sets (forewing mKO mutants and corresponding contralateral WT) are shown with one of the wings horizontally flipped to show the mutant wings in left-to-right orientation.. Cyan arrows : small mutant clones. Cyan asterisks : large mutant clones.

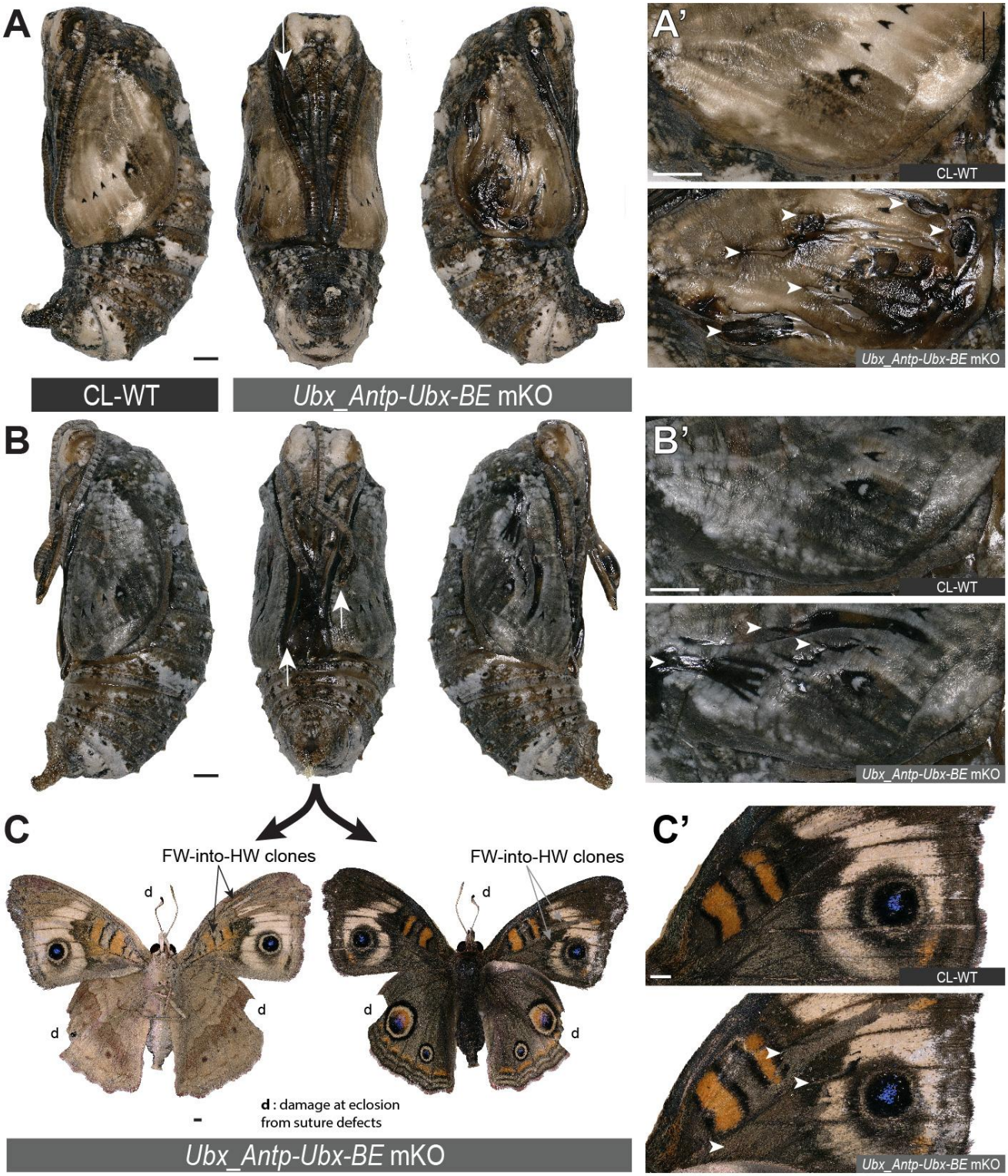


Figure S3. Pupal defects following FW→HW homeosis in *Antp-Ubx_BE* crispants. (**A-B**) Contralateral (CL) and, forewing mosaic knockout (mKO) mutants following CRISPR targeting of *Antp-Ubx_BE* in *J. coenia*. The two pupae show suture defects in the midline appendages (arrows). (**A'-B'**) Magnified views of the crispant forewings, showing defective cuticle (arrowheads). (**C-C'**) Crispant adult butterfly emerged from the pupa in panel B. White arrowheads in C' highlight the match between dorsal forewing clones and the pupal forewing cuticle defects shown in B'. Scale bars : 1 mm.

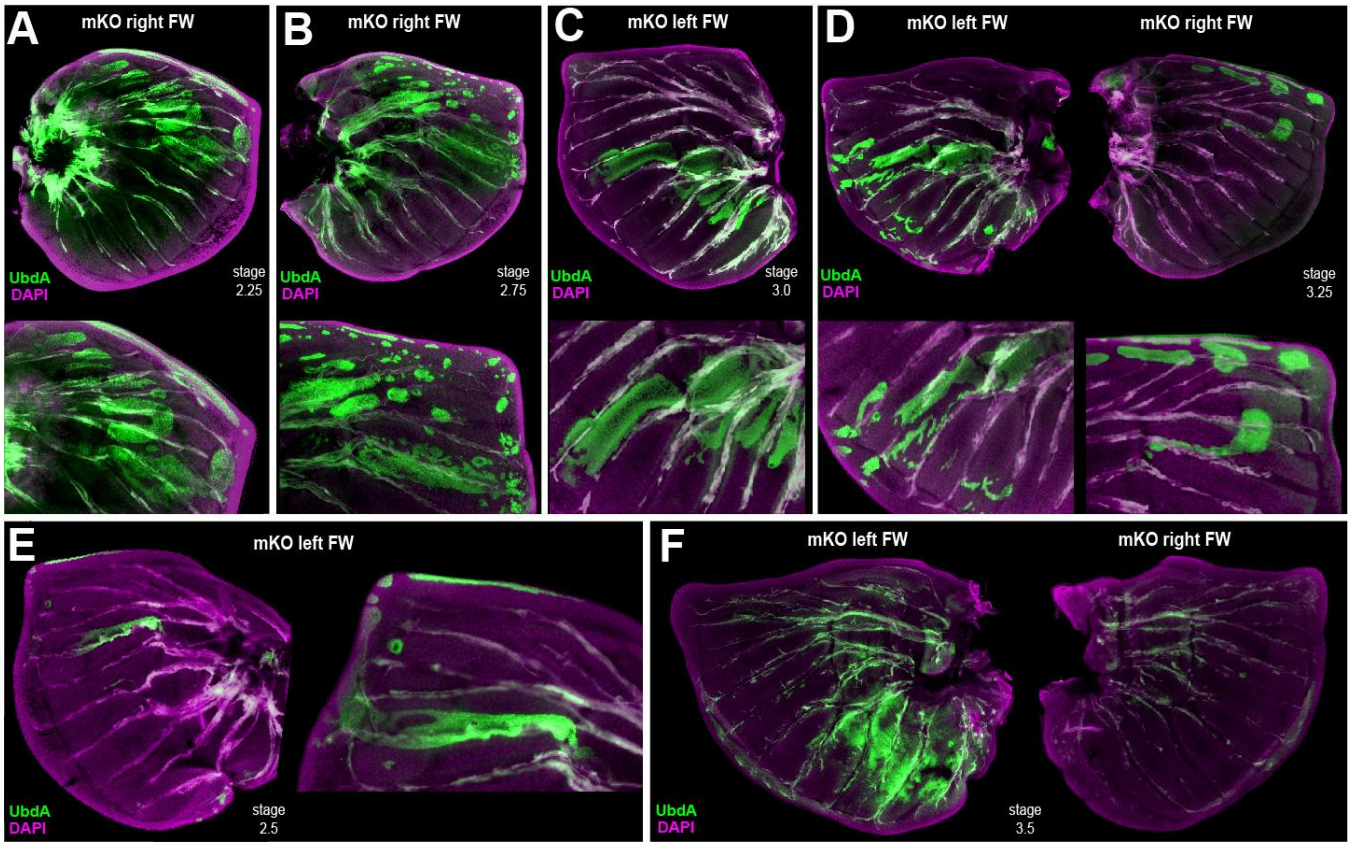


Figure S4. Additional examples of ectopic UbdA and FW→HW homeosis in *Antp-Ubx_BE* crispant larval forewings. (**A-F**) Each panel shows forewings with ectopic detection of UbdA (FP6.87 monoclonal antibody, green), dissected at the fifth instar stage. Panels D and F are wing sets from individual crispants. Panels E and C are mutant contralateral wings of the mutant forewings shown in Figs. 4D and E, respectively.

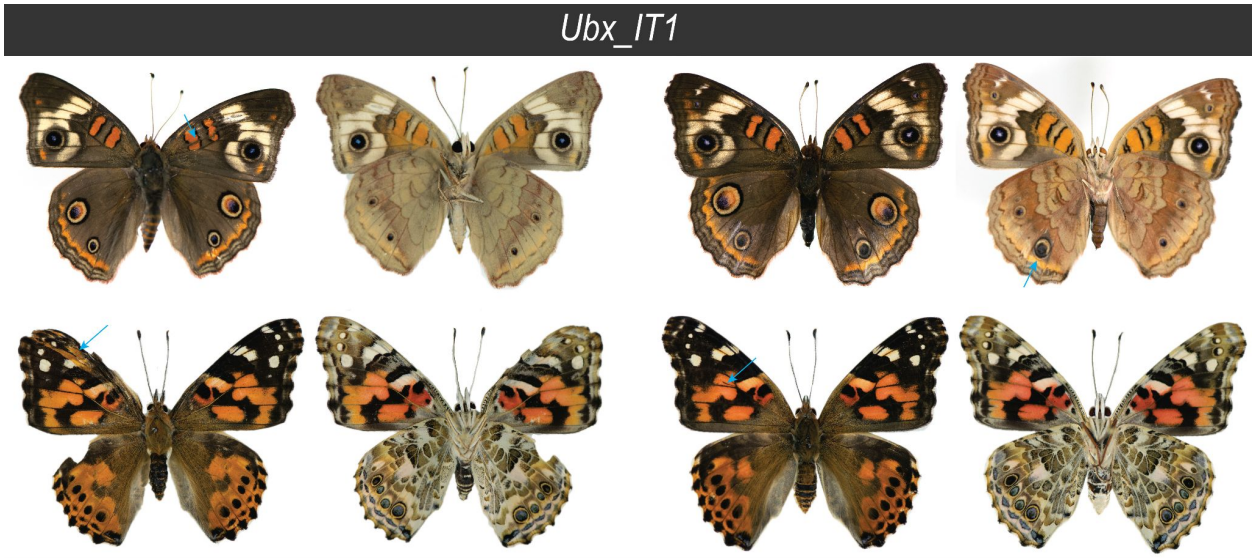


Figure S5. Additional mutant phenotypes from CRISPR-mediated interrogation of *lncRNA_Ubx-IT1* 5' region in *J. coenia* (top) and *V. cardui* (bottom). Cyan arrows : mutant clones.

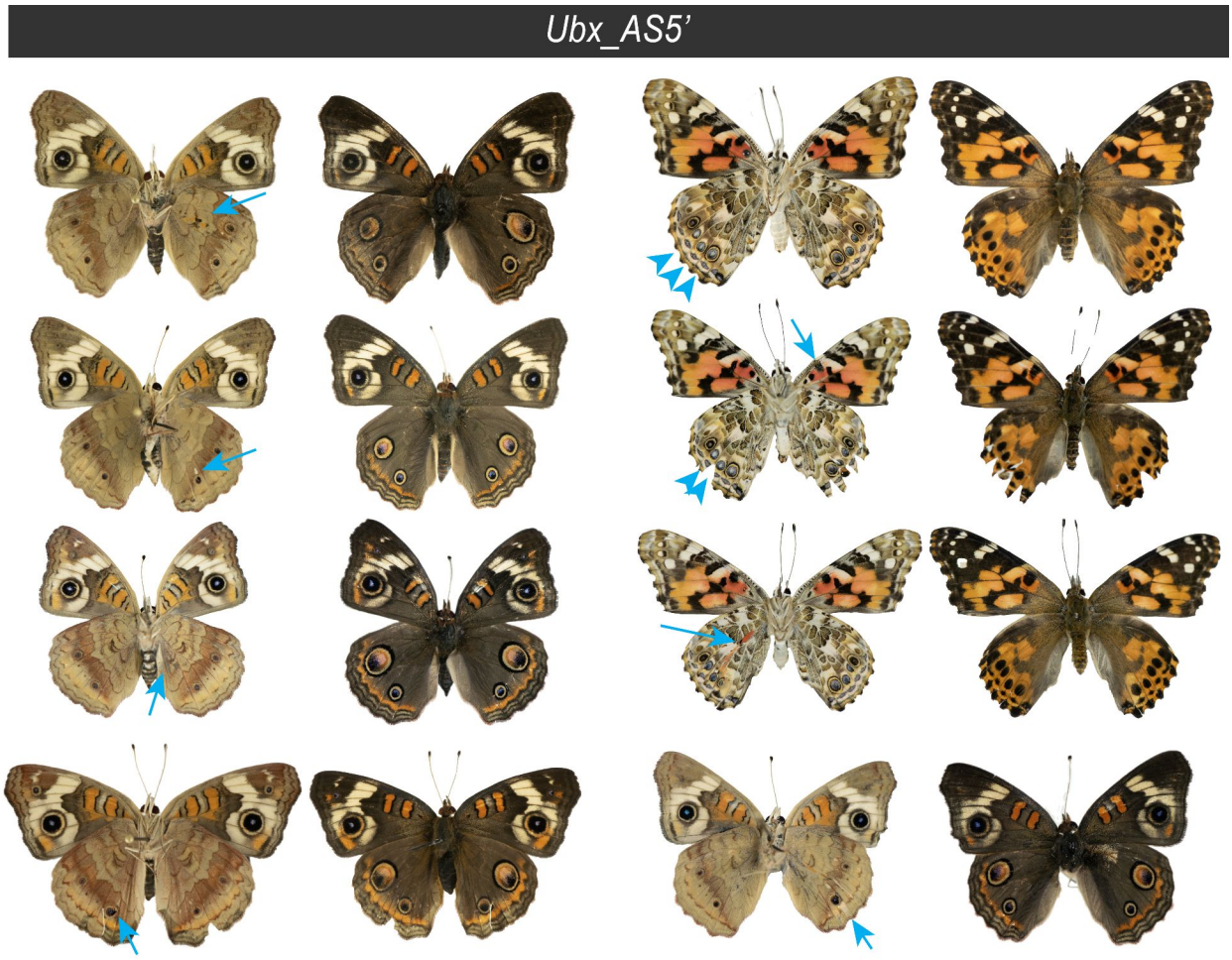


Figure S6. Additional mutant phenotypes from CRISPR-mediated interrogation of the *lncRNA_Ubx-AS*5' region in *J. coenia* and *V. cardui*. Cyan arrows : mutant clones. Cyan arrowheads : white eyespot foci.

Figure S7. Additional mutant phenotypes from CRISPR-mediated interrogation of *CRM11* in *J. coenia* and *V. cardui* show bidirectional homeoses and non-homeotic eyespot changes. Cyan arrows : mutant clones. Cyan arrowheads : white eyespot foci.

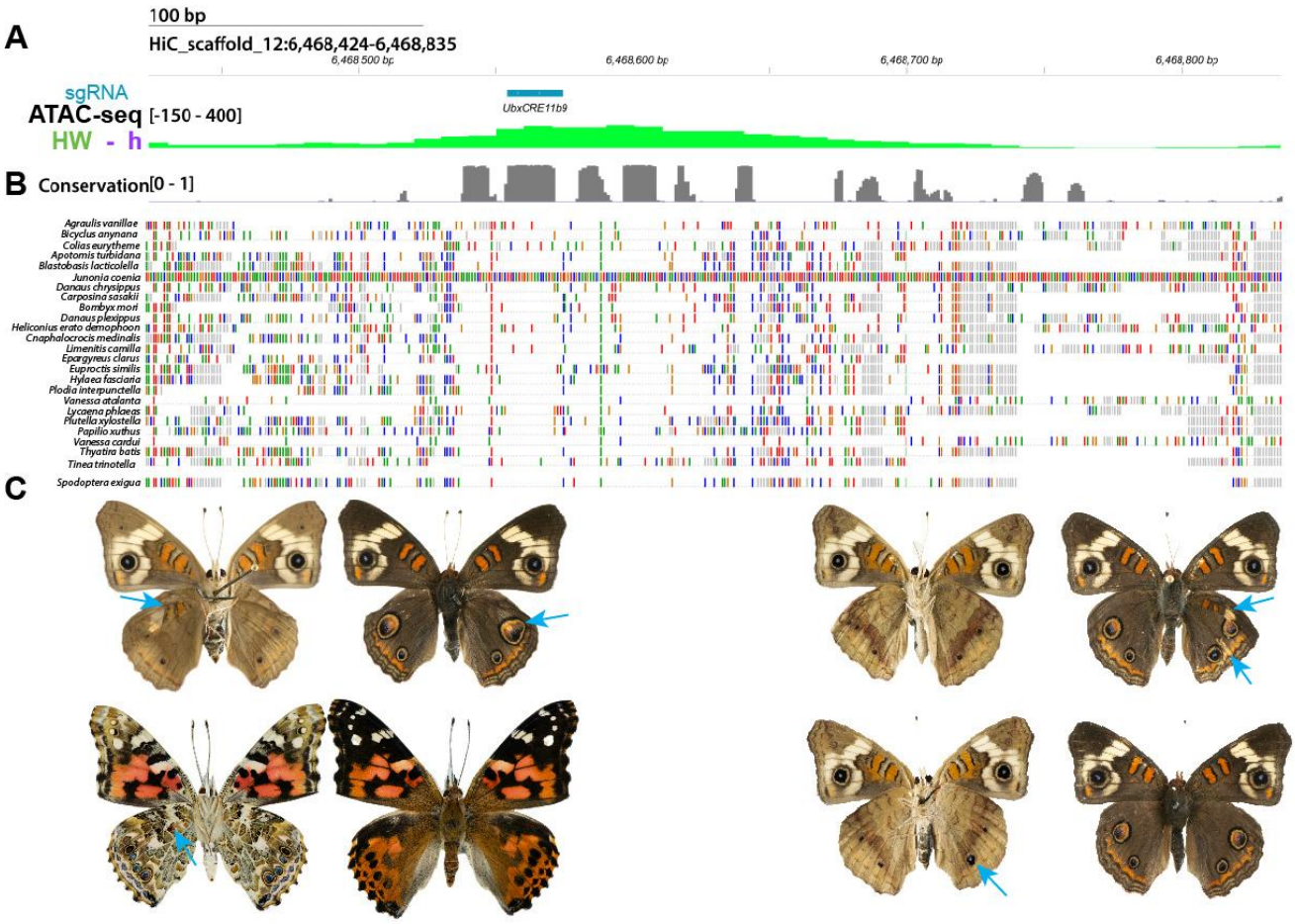
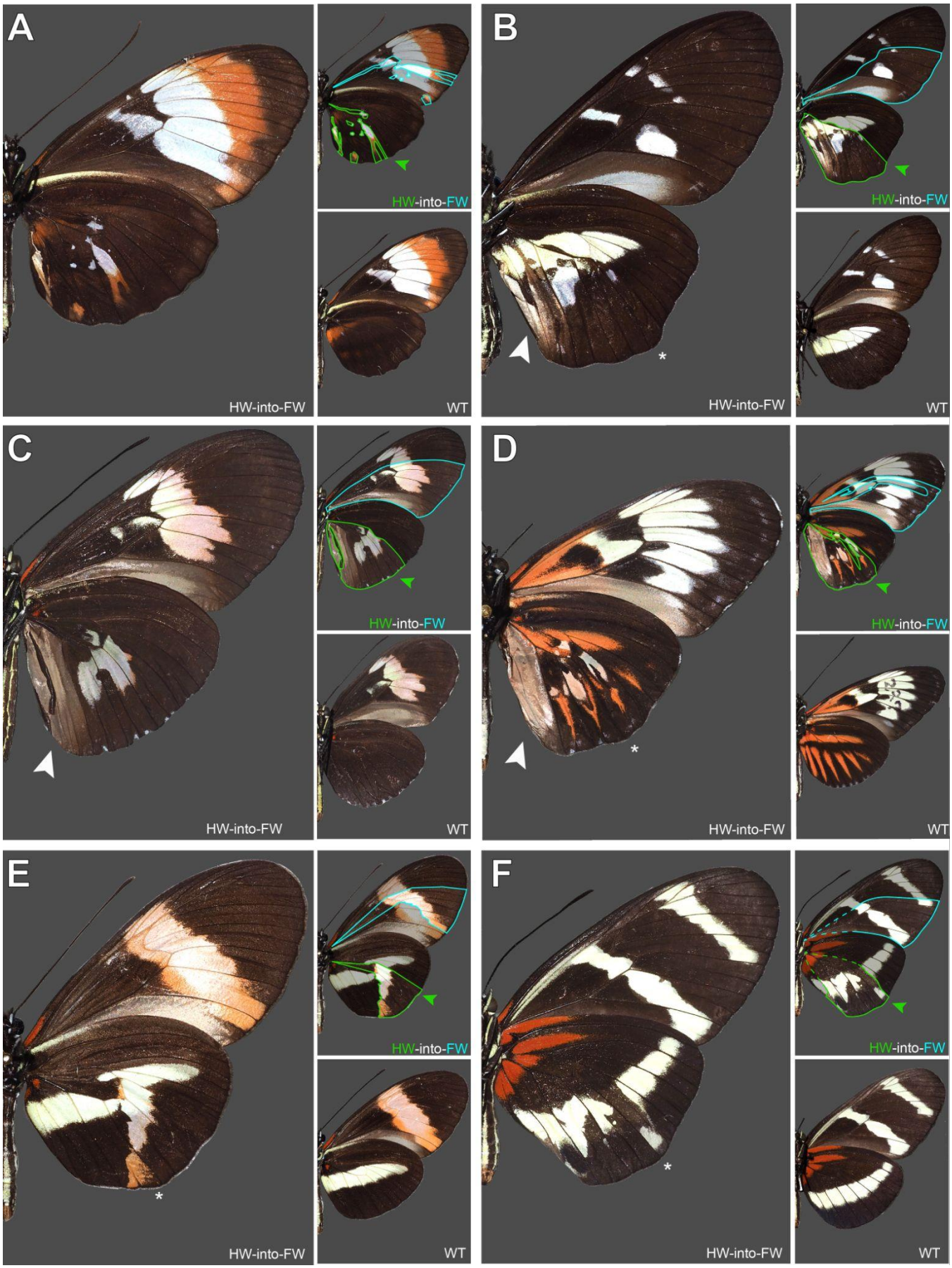


Figure S8. CRISPR perturbation of the conserved *Ubx_CRE11b* results in HW→FW homeoses. (**A-B**) The *UbxCRE11b9* sgRNA targets a hindwing-enriched ATAC peak with strong conservation across genomes from 23 Lepidoptera and 2 Trichoptera species (gray : PhastCons scores). Colored bars denote variation from the *J. coenia* reference (**C**) *Jc_UbxCRE_11b9* crispant butterflies exclusively showed HW→FW transformed clones (cyan arrows in both *J. coenia* and *V. cardui*).


A
HW-into-FW
HW-into-FW
WT
B
HW-into-FW
HW-into-FW
WT
C
HW-into-FW
HW-into-FW
WT
D
HW-into-FW
HW-into-FW
WT
E
HW-into-FW
HW-into-FW
WT
F
HW-into-FW
HW-into-FW
WT

820

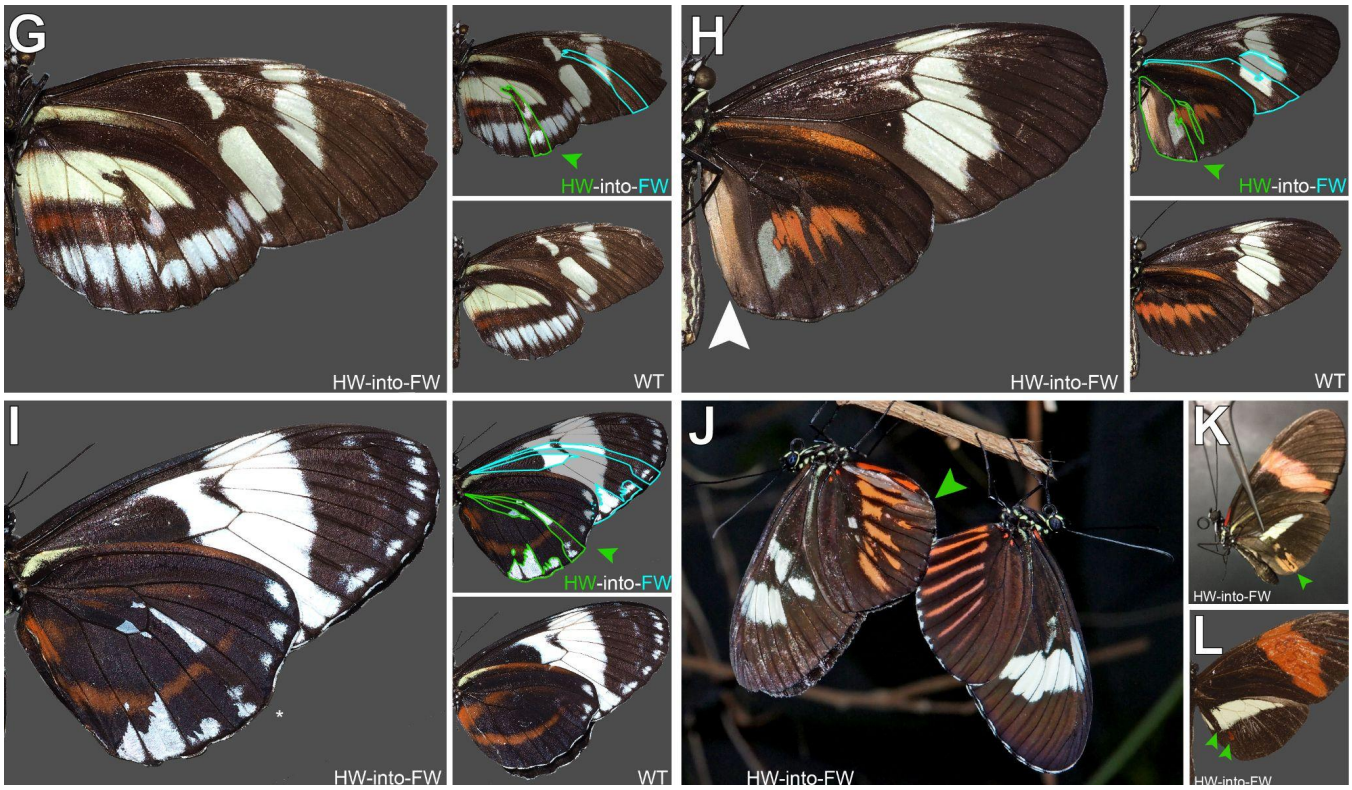


Figure S9 (previous page, continued above). Hindwing homeoses in *Heliconius* butterfly spontaneous mutants from pure stocks, hybrid cultures and wild-caught individuals from the L.E. Gilbert collection (UT Austin). White arrowheads: homeotic clones including the acquisition of ventral forewing coupling scales. Asterisks : local deformation of hindwings relative to wild-type. All hindwing homeoses are ventral except in panel L. **A.** *Heliconius cydno galanthus x H. melpomene rosina* (Costa Rica), cross J31, August 1987. **B.** *Heliconius cydno gustavi*, captive stock from Saladito (Colombia), September 1991. **C.** *Heliconius melpomene madeira* (Brazil) *x Heliconius melpomene plesseni* (Ecuador), September 2012. **D.** *H. m. rosina* (Costa Rica) *x Heliconius melpomene madeira* (Brazil) *x H. cydno galanthus* (Costa Rica) mixed population, December 2015. **E.** *H. m. rosina,* captive stock from Osa Peninsula (Costa Rica), September 1991. **F.** *Heliconius hewitsoni,* captive stock from Osa Peninsula (Costa Rica), July 2005. **G.** *Heliconius cydno cydnides*, captive stock from natural hybrid zone in Dagua Pass (Colombia), May 1989. **H.** *H. m. rosina* (Costa Rica) *x H. m. madeira* (Brazil) *x H. c. galanthus* (Costa Rica) mixed population, June 2016. **I.** *H. c. galanthus x H. m. rosina* crossed three times, and back to *H. c. galanthus*, August 2014. **J.** *Heliconius melpomene malleti* (Ecuador) x *H. m. plesseni* (Ecuador) hybrid stock, 2010. **K.** *H. m. rosina* captive stock, Costa Rica. **L.** *H. m. rosina* captive stock, Osa Peninsula (Costa Rica), March 1987, in dorsal view.