

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

Amruta Tendolkar, Anyi Mazo-Vargas, Luca Livraghi, Joseph J. Hanly, Kelsey C. Van Horne, Lawrence E. Gilbert, Arnaud Martin 

Department of Biological Sciences, The George Washington University, Washington, DC, United States • Smithsonian Tropical Research Institute, Panama • Department of Integrative Biology, University of Texas – Austin, Austin, TX, United States

Reviewed Preprint

Revised by authors after peer review.

About eLife's process

Reviewed preprint version 2
January 10, 2024 (this version)

Reviewed preprint version 1
October 13, 2023

Posted to preprint server
August 18, 2023

Sent for peer review
August 7, 2023

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

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.

eLife assessment

This **valuable** paper examines the Bithorax complex in several butterfly species, in which the complex is contiguous and not split, as it is in the well-studied fruit fly *Drosophila*. Based on genetic screens and genetic manipulations of a boundary element involved in segment-specific regulation of *Ubx*, the authors provide **convincing** evidence for their conclusions, which could be strengthened by additional data and analyses in the future. The data presented are relevant for those interested in the evolution and function of Hox genes and of gene regulation in general.

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 (Lewis 1978 [↗](#)). 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 [↗](#)). For instance, the deletion of a small region situated between *Ubx* and *abd-A* produces the *Front-ultraabdominal* phenotype (*Fub*) where the first abdominal segment (A1) is transformed into a copy of the second abdominal segment A2, due to a gain-of-expression of *abd-A* in A1 where it is normally repressed (Bender and Lucas 2013). At the molecular level, the *Fub* boundary is enforced by insulating factors that separate Topologically Associating Domains (TADs) of open-chromatin, while also allowing interactions of *Ubx* and *abd-A* enhancers with their target promoters (Postika et al. 2018 [↗](#); Srinivasan and Mishra 2020 [↗](#)). Likewise, the *Fab-7* deletion, which removes a TAD boundary insulating *abd-A* and *Abd-B* (Moniot-Perron et al. 2023 [↗](#)), transforms parasegment 11 into parasegment 12 due to an anterior gain-of-expression of *Abd-B* (Gyurkovics et al. 1990). By extrapolation, one may expect that if the *Drosophila* *Hox* locus was not dislocated into two complexes, *Antp* and *Ubx* 3D contact domains would be separated by a Boundary Element (BE), and that deletions similar with *Fub* and *Fab-7* mutations would result in gain-of-function mutations of *Ubx* that could effectively transform T2 regions into T3 identities.

The BX-C locus also includes non-coding RNAs (Pease et al. 2013), some of which are processed into miRNAs known to repress *Ubx* and *abd-A* (Gummalla et al. 2012; 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 (Petruk et al. 2006; Ibragimov et al. 2023). 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 insect *Hox* genes are regulated outside of *Drosophila*, 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 (mKO) 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). A 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 *Aglaia* (*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 *lncRNA: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.

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. 2023; 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

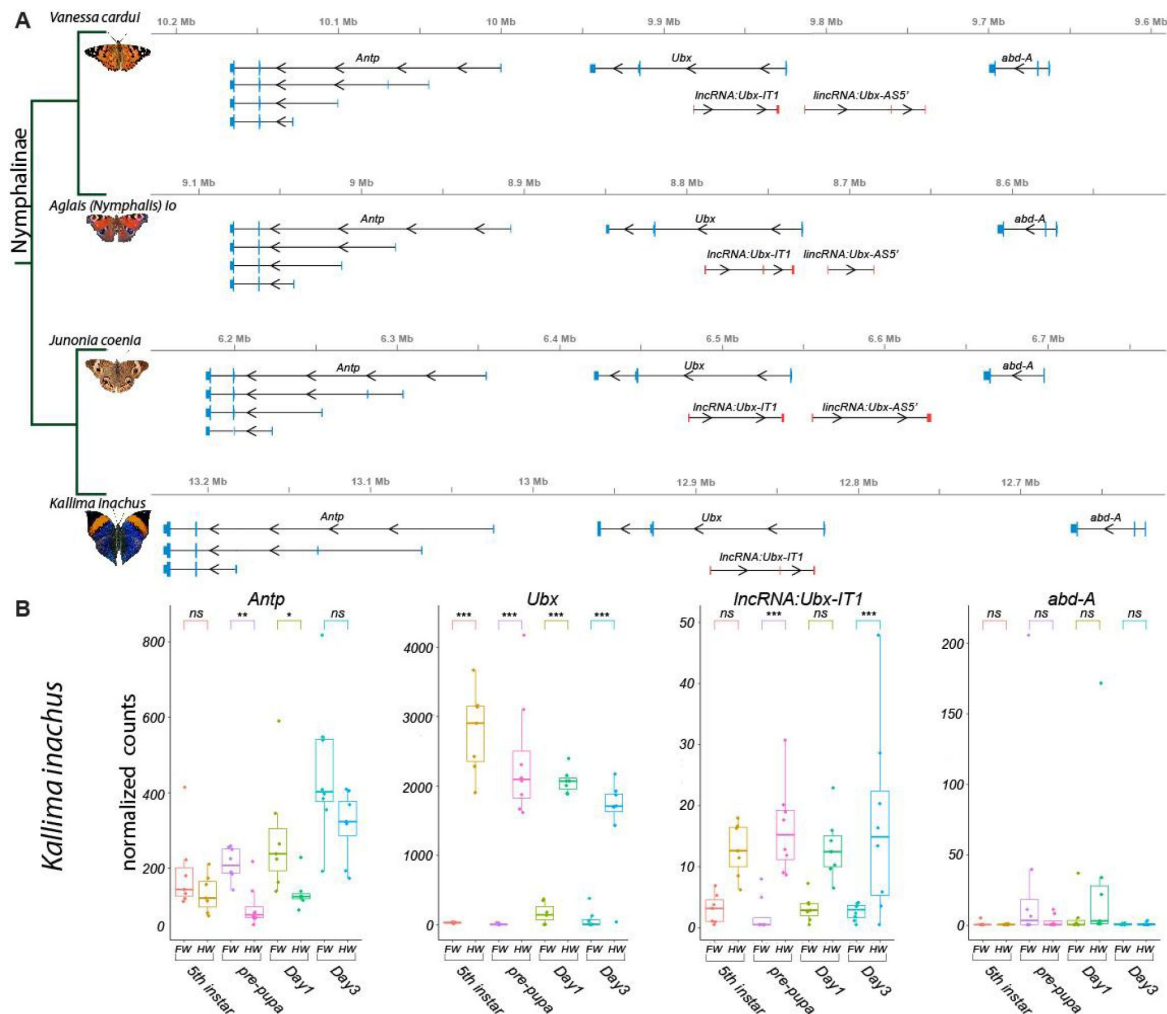


Figure 1.

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; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$.

TAD insulating role in mediating chromatin loop-extrusion (Guo et al. 2015 [↗](#)). This concordance between Hi-C profiling and CTCF motif prediction thus suggests that *Antp-Ubx_BE* region might function 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* collinear 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 differential 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 [↗](#)).

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 *Ubx* to be insulated from the 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.

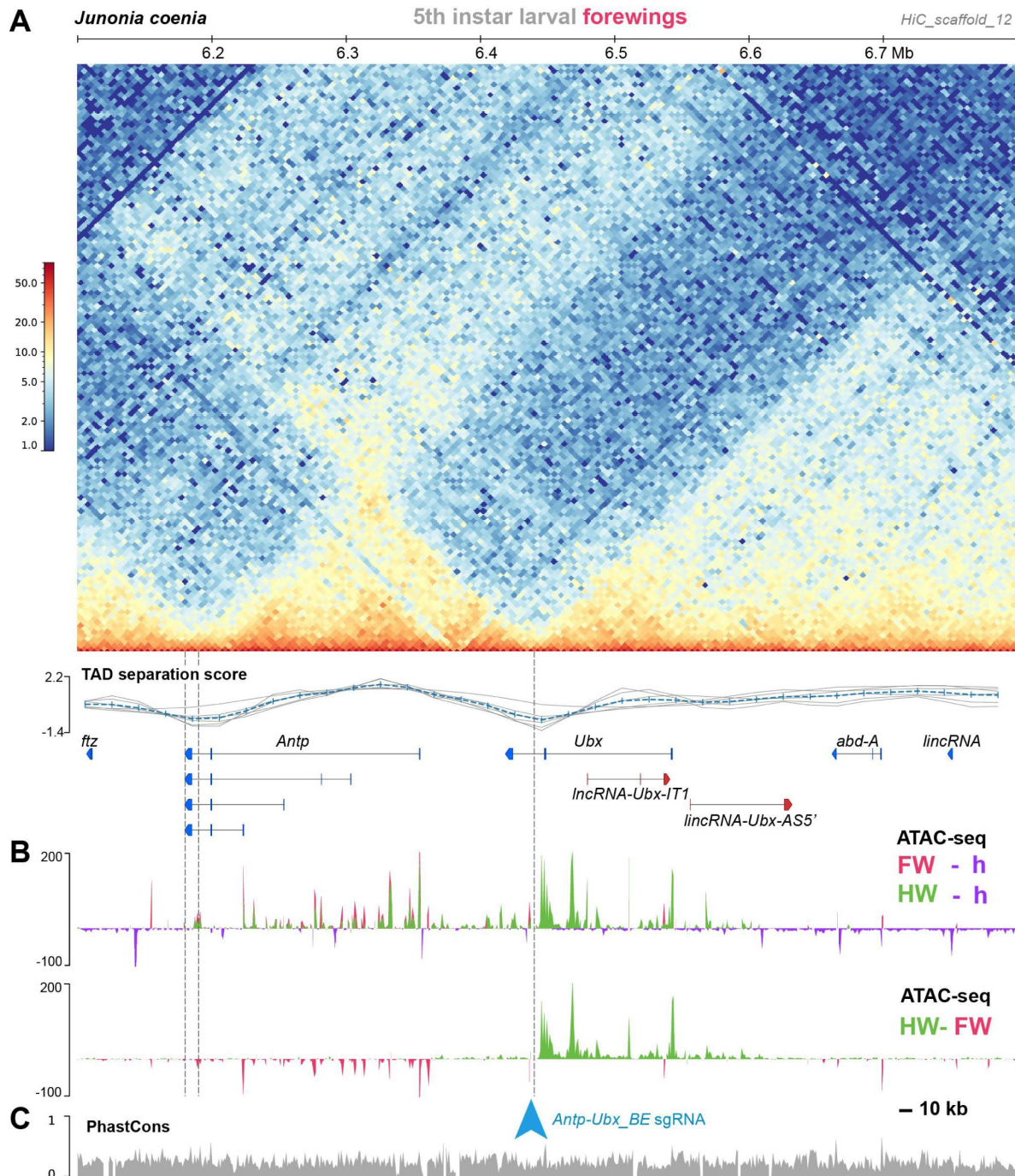


Figure 2.

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.

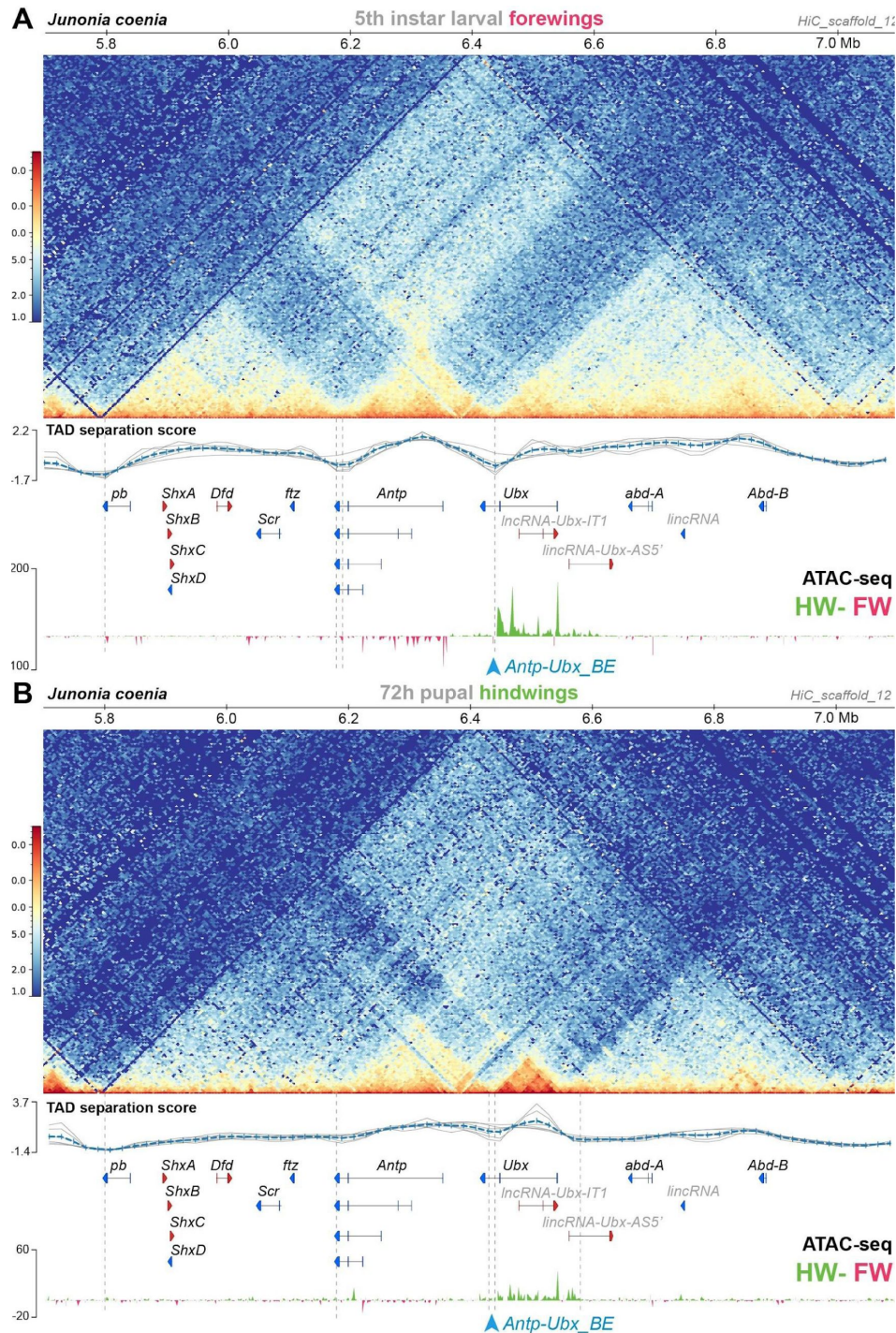


Figure 3.

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.

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).

Homeotic clones are first visible in *Antp-Ubx_{BE}* crispants at the pupal stage, with streaks of thinner cuticle, sometimes associated with local necrosis or with suture defects in the ventral midline, in particular where leg and wing pouches adjoin (**Fig. S3**). When dissected out of the pupa, these mutant forewings also show streaks of more transparent wing epithelium, concomitantly with the territories of thinner cuticle above, and PCR genotyping of affected pupal wing tissue confirmed the presence of CRISPR-induced mutations at the *Antp-Ubx_{BE}* target site (**Fig. S4**). Color pattern homeotic clones were salient at the adult stage, with for example, clonal losses of the forewing specific white-band, and partial acquisitions of the large M_1 - M_2 hindwing eyespot. In one specimen, an ectopic, partial M_1 - M_2 hindwing eyespot appeared in the R_5 - M_1 region, suggesting a perturbation of the eyespot induction process in this wing. Nymphalid forewings have five radial veins (R_{1-5}), which provide sturdiness for flight (Wootton 1993), while hindwings have only two Radial veins. Forewing homeotic mutants showed mosaic venation defects in the Radial vein area (**Fig. 4B**). Finally, higher expressivity mutant forewings were smaller and rounder, reminiscent of hindwing shape.

Next, we dissected fifth instar larval wing disks from developing *Antp-Ubx_{BE}* crispants, and monitored the expression of Ubd-A (Ubx and Abd-A epitopes), normally restricted to the hindwing and only present in the forewing peripodial membrane (Weatherbee et al. 1999). Crispants showed forewing clones with strong ectopic expression of Ubd-A (**Figs. 4D-E** and **S5**). This result supports the inference that *Antp-Ubx_{BE}* forewing homeoses are due to the de-repression of *Ubx* in this tissue.

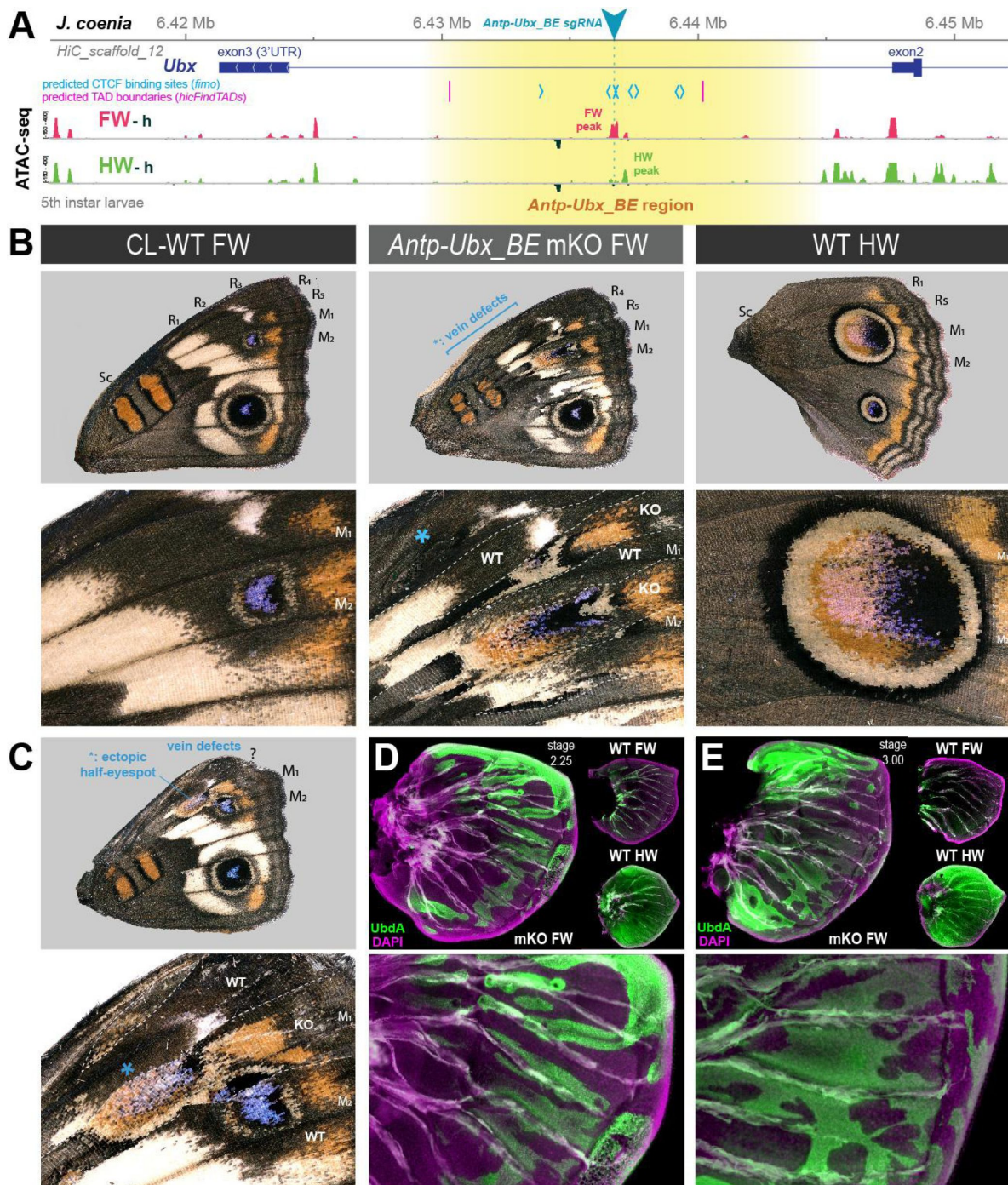


Figure 4.

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.

Mutational interrogation of lncRNA-encoding regions at the *Ubx* locus

We used CRISPR mutagenesis to test the function of the two lncRNA-encoding loci at the *Ubx* locus. Mutagenesis of the *Ubx-IT1* first exon in *J. coenia*, and of the *Ubx-IT1* promoter in *V. cardui*, both resulted in crispants with small homeotic phenotypes in forewings and hindwings (Figs. 5 and S6). This result contrasts with *Ubx* exon mKO experiments, which only generate hindwing phenotypes (Tendolkar et al. 2021). Given the scarcity of *Ubx-IT1* crispants obtained (11 out of 236 adults), and the small size of the homeotic clones within them, we infer the occasional phenotypes may be due to rare alleles. Thus, rather than evidence of functionality of the *Ubx-IT1* transcript, the homeotic phenotypes may rather reflect the effects of regulatory perturbation on *Ubx* itself, with some random mutations in this intronic region resulting in hindwing *Ubx* loss-of-function, and some others triggering derepression in forewings. Likewise, next we mutagenized the first exon of *Ubx-ASS'*, located upstream of the *Ubx* promoter, and obtained twelve hindwing mutants and a single forewing mutant (Fig. 6 and S7). As with *Ubx-IT1* CRISPR experiments, these results may be explained by regulatory disruption of *Ubx* transcription, with a higher ratio of hindwing phenotypes compared to forewings linked to the proximity of the *Ubx* promoter. Overall, we conclude that the mutational interrogation at these loci can result in dual loss (hindwing) and gain (forewing) of *Ubx* function effects. Deciphering whether or when these effects affected *Ubx* expression via local *cis*-regulatory modules, impairment of lncRNA transcripts, or larger indels overlapping with *Ubx* exons, will require further study (see Discussion).

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 subdivided 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 single target in 11c (*UbxCRE11c5* in *V. cardui*) yielded dual homeoses : FW→HW and HW→FW (Figs. 7B-D and S8). Hindwing effects were reminiscent of *Ubx* protein coding knockouts (Tendolkar et al. 2021), indicating that these crispant alleles with a hindwing phenotype produce *Ubx* loss-of-function effects. Individuals with hindwing clones were 2.75 times more common than individuals with forewings in this dataset. Similarly to the lncRNA loci perturbation experiments, dual homeoses may indicate the presence of hindwing activators and forewing repressors in the *CRM11* region, with various CRISPR alleles producing a spectrum of indels and effects (see Discussion). It is noteworthy that while single-target experiments showed little lethality (55% hatching rate), dual or quadruple injection mixes resulted in low hatching rates of injected embryos (~ 10%). Multiple targeting thus appears to induce high-rates of embryonic lethality, possibly due to chromosomal damage (Cullot et al. 2019; Zuccaro et al. 2020). Dual targeting with *a2+c5* also yielded partial HW→FW homeoses in *V. cardui* under the form of ectopic white eyespot foci phenotypes (Fig. 7E), as occasionally observed in *Ubx* null crispants (Tendolkar et al. 2021), seemingly due to hypomorphic or heterozygous allelic states.

Next, we focused on a single target shared between both *V. cardui* and *J. coenia* in the 11b sub-domain. A whole genome alignment between 23 lepidopteran species and 2 trichopteran species indicated that region 11b is deeply conserved, suggesting important functional constraints on its sequence (Fig. S9A-B). Mutagenesis of 11b using a single guide RNA (*Ubx11b9*) yielded a relatively high hatching rate (mean = 51.8 %), indicating a rare occurrence of the deleterious mutational effects observed in multiple targeting (see above). Four *J. coenia* crispants and two *V. cardui* crispants were obtained, all exclusively showing hindwing phenotype devoid of forewing effects. HW→FW homeoses included a variety of phenotypes all seen in *Ubx* CDS mutants (Tendolkar et al. 2021), including transformations of the orange Discalis elements and the white band in *J. coenia*,

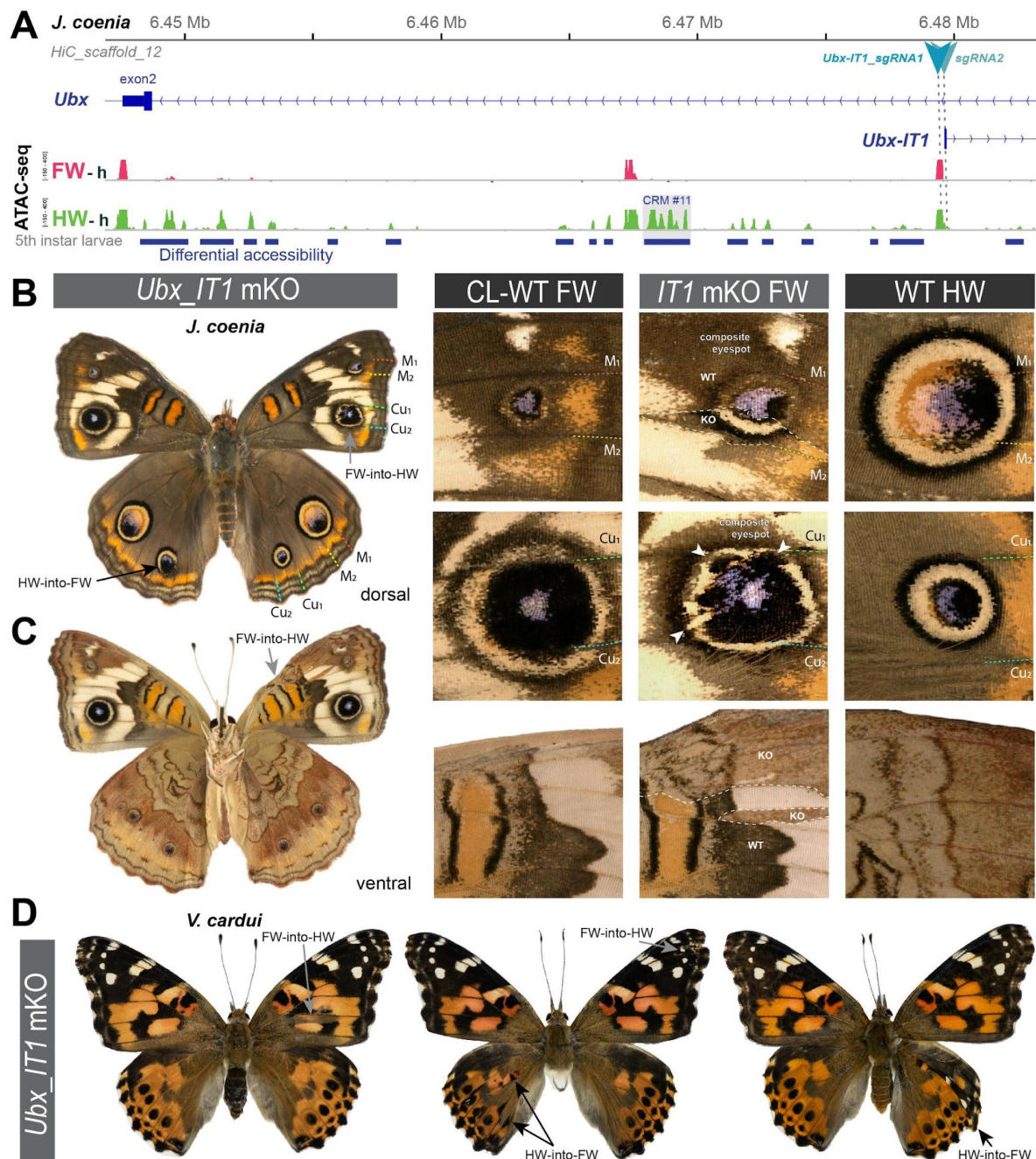


Figure 5.

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*.

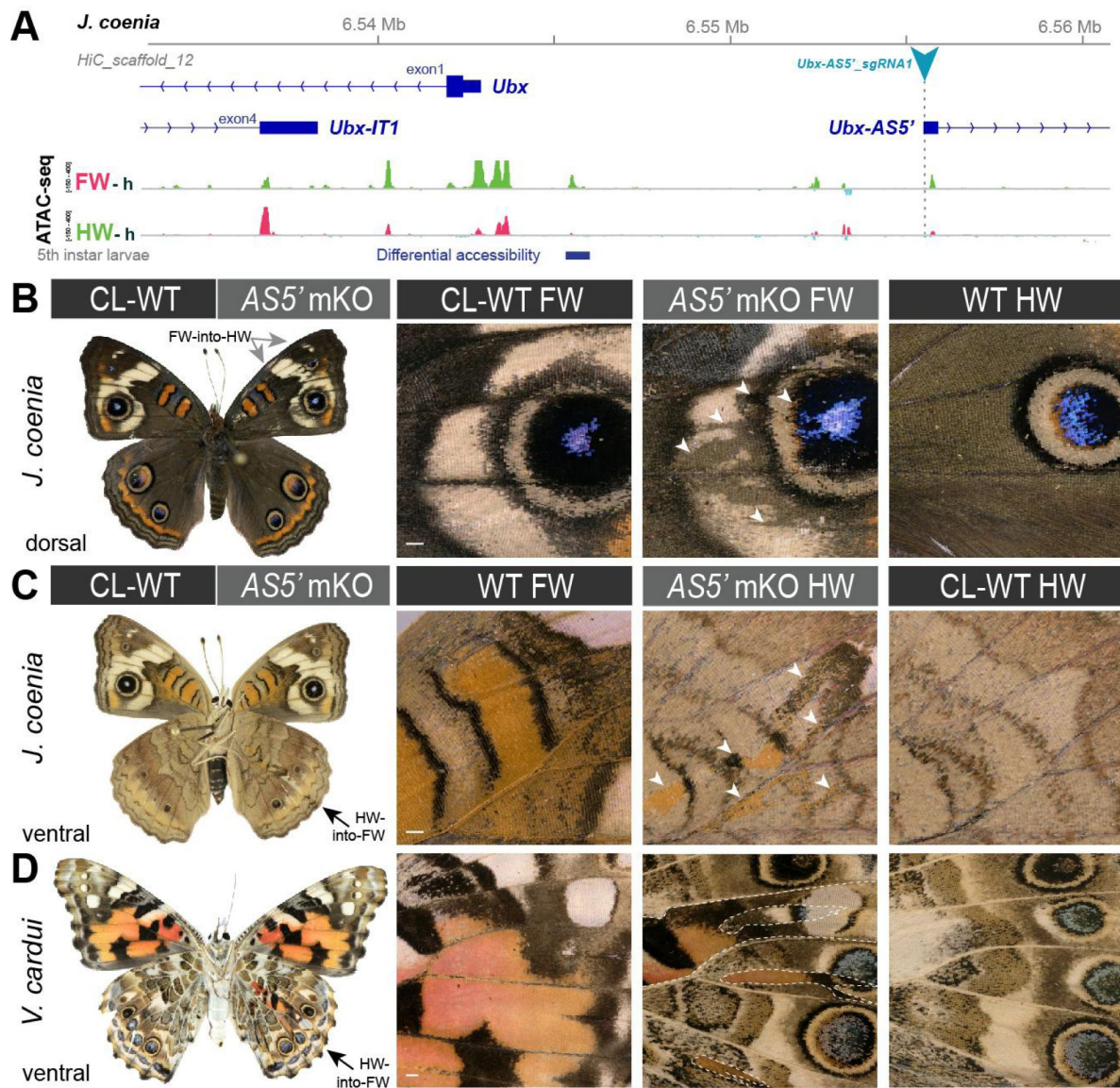


Figure 6.

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 μ m.

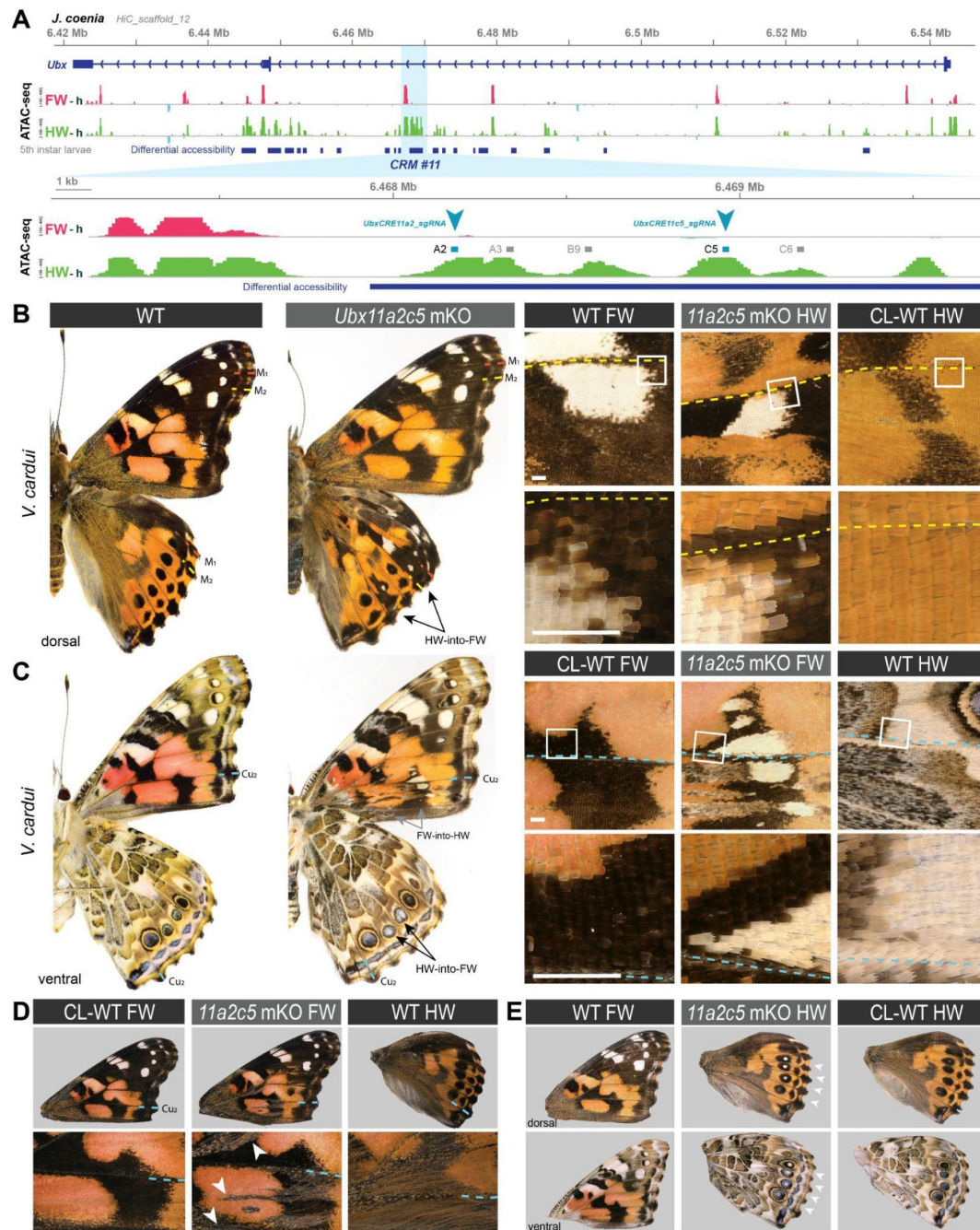


Figure 7.

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 *Ubx*CRE11a2c5, including homeoses in color patterns and scale morphology. (D) Additional example of a *V. cardui* *Ubx*CRE11a2c5 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 μ m.

and partial shifts in eyespot identity (Fig. S9C). Together the consistency in direction of transformations and the deep conservation of the *11b* region suggests it may encode an enhancer necessary for the transcriptional activation of *Ubx* in hindwings.

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. S10), 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. S10B-D, H). Homeoses also include noticeable local changes in wing shape, particularly in hindwings (asterisks in Fig. S10). 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.

Discussion

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 [↗](#)). Ideally, regulatory CRISPR-induced alleles should require genotyping in a second (G_1) generation to be properly matched to a phenotype (Mazo-Vargas et al. 2022 [↗](#)). Possibly because of lethal effects, we failed to pass *Ubx* locus G_0 mutations to a G_1 generation for genotyping, and were thus limited here to mosaic analysis. As adult wings scales are dead cells, the genetic material building a given color phenotype is lost at this stage, but we circumvented this issue by genotyping a pupal forewing displaying an homeotic phenotype in the more efficient *Antp-Ubx_{BE}* perturbation experiment (Fig. S4). In this case, PCR amplification of a 600 bp fragment followed by Sanger sequencing recovered signatures of indel variants, with mixed chromatograms starting at the targeted sites. But in all other experiments (*CRM11*, *IT1*, and *AS5'* targets), we did not genotype mutant tissues, as they were only detected in adult stages and generally with small clone sizes. Some of these clones may have been the results of large structural variants, as 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 [↗](#)), and we are unable to assign a specific genotype to the homeotic effects we obtained in *CRM11*, *IT1* and *AS5'* perturbation assays.

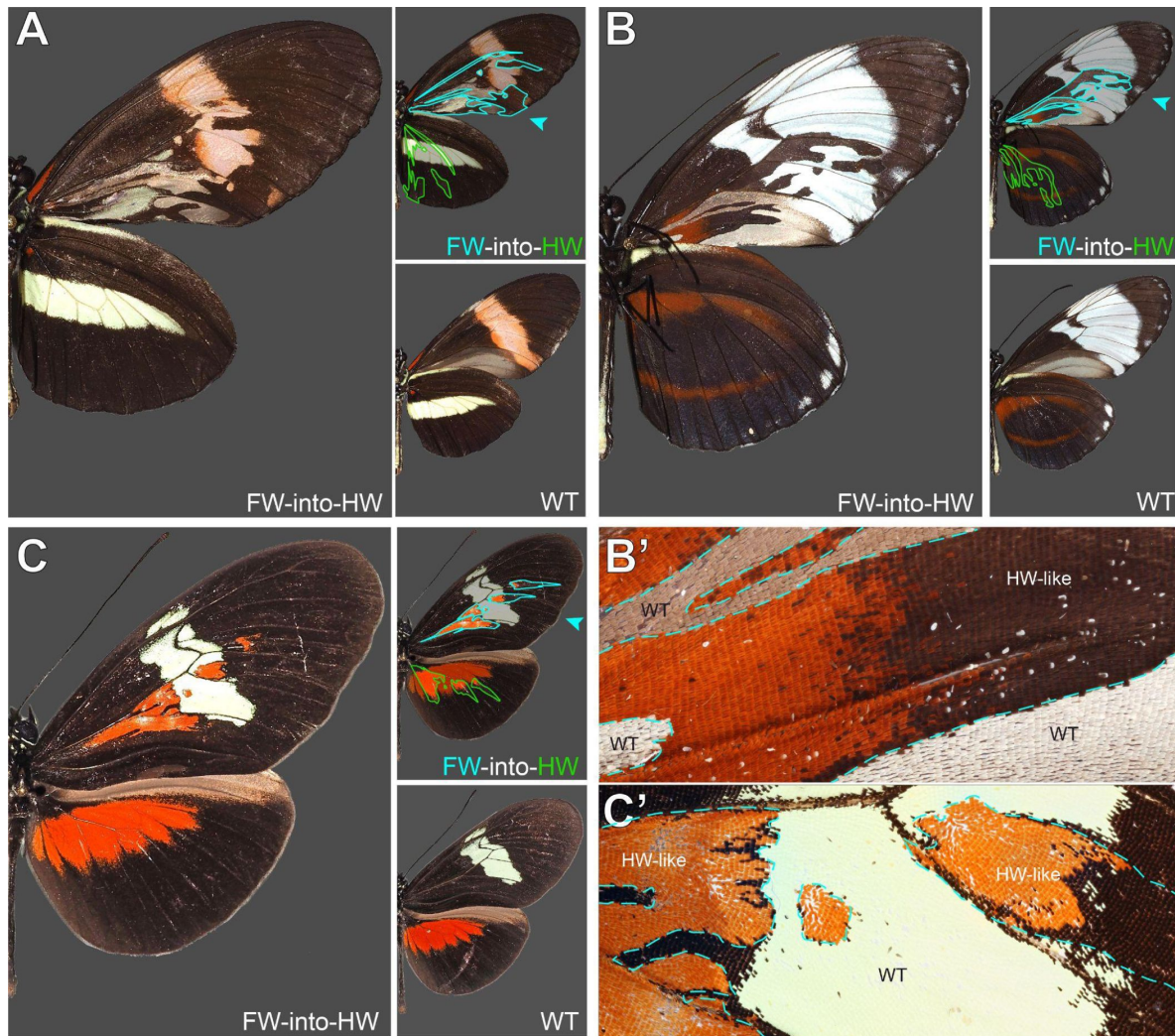


Figure 8.

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.

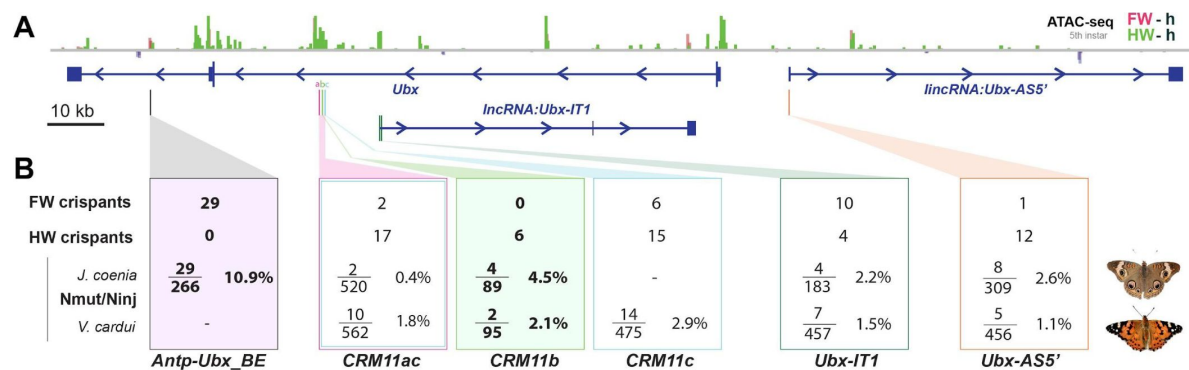


Figure 9.

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.

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
<i>J. coenia</i>	<i>Antp-Ubx_BE</i>	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%
<i>V. cardui</i>	<i>IT1_sgRNA1</i>	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%
<i>J. coenia</i>	<i>IT1_sgRNA2</i>	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%
<i>V. cardui</i>	<i>AS5_sgRNA1</i>	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%
<i>J. coenia</i>	<i>AS5_sgRNA1</i>	309	181	181	181	8	2-4.5	500 : 250	58.6%	2.6%
<i>J. coenia</i>	<i>Ubx11a2+3 +c5+6</i>	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%
<i>V. cardui</i>	<i>Ubx11a2+c5</i>	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%
<i>V. cardui</i>	<i>Ubx11c5</i>	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%
<i>V. cardui</i>	<i>Ubx11b9</i>	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%
<i>J. coenia</i>	<i>Ubx11b9</i>	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 instars larvae that were dissected for immunostainings, of which 7 were mutants (as evidenced by ectopic Ubx in forewings), and 3 dissected mutant pupae

Table 1.

CRISPR mutational interrogation experiments at putative *Ubx* regulatory regions

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*. As 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. Of note, a systematic in-situ hybridization survey (Pease et al. 2013 [↗](#)) showed that *Drosophila* embryos express an antisense transcript in its 5' region (*lncRNA:bxd*), as well as within its first intron (*lncRNA:PS4*). It is thought that *Drosophila bxd* regulates *Ubx*, possibly by transcriptional interference or by facilitation of the *Fub-1* boundary effect (Petruk et al. 2006 [↗](#); Ibragimov et al. 2023 [↗](#)), while the possible regulatory roles of *PS4* remain debated (Hermann et al. 2022 [↗](#)). While these dipteran non-coding transcripts lack detectable sequence similarity with the lepidopteran *IT1* and *AS5'* transcripts, further comparative genomics analyses of the *Ubx* region across the holometabolan insect phylogeny should clarify the extent to which Hox cluster lncRNAs have been conserved or independently evolved.

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 loss 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 deletion alleles resulting in a TAD fusion and misexpression effect 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. Lastly, it is worth noting that the *Antp/Ubx* TAD boundary we identified is intragenic, within the last intron of *Ubx*. It is unclear if this feature affects transcription, but this configuration might be analogue to the *Notch* locus in *Drosophila*, which includes a functional TAD boundary in an intronic position (Arzate-Mejía et al. 2020 [↗](#)).

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. S10). 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 *ilAgIlox1.1*; *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 preliminary gene models provided by Peiwen Yang and Wei Zhang (Wang et al. 2022 [↗](#)). All transcripts analyses were performed using the *STAR* intron-aware aligner and *DESeq2* expression analysis package as previously described (Love et al. 2014 [↗](#); Dobin and Gingeras 2016 [↗](#); Hanly et al. 2019 [↗](#), 2022 [↗](#)). Expression levels were calculated as genome-wide normalized counts and pairwise Wald tests were performed to assess differential expression between forewings and hindwings. Multiple test adjustment was performed using Benjamini and Hochberg correction.

Hi-C and ATAC-seq analyses

Hi-C data from *J. coenia* fifth instar larval forewings and 48-72 h APF pupa hindwings are available at the NCBI SRA BioProject *PRJNA641138* (van der Burg et al. 2020 [↗](#)). Triplicated ATAC-seq datasets for larval and pupal wing and head tissues of *J. coenia* and *V. cardui* (van der Burg et al. 2019 [↗](#), 2020; Mazo-Vargas et al. 2022 [↗](#)) are available on the NCBI SRA BioProjects *PRJNA497878*, *PRJNA695303*, and *PRJNA559165*. All the ATAC-seq and Hi-C data were re-analysed on *J. coenia* and *V. cardui Ubx* genomic regions as previously described (Mazo-Vargas et al. 2022 [↗](#)). Briefly, matrices of interactions were constructed by mapping paired reads against the *Junonia coenia* genome (Mazo-Vargas et al., 2022 [↗](#)) using *hicBuildMatrix* (Ramírez et al. 2018 [↗](#)). Matrices from larvae and pupae were normalized using *hicNormalize* and corrected with the Knight-Ruiz matrix balancing algorithm. The definitions of topologically associating domains (TADs) can be influenced by various factors such as the choice of software, parameters, sequencing depth, and the presence of experimental noise. To ensure reliability, it is recommended to compare TAD calls with independent datasets, such as histone marks or known factors associated with TAD boundaries. In the absence of these specific datasets, we employed a different combination of parameters in the *hicFindTADs* tool and compared the resulting TAD calls. HiC matrices at 10 kb and 20 kb bin resolutions were utilized, and TAD insulation scores were evaluated using a false-discovery rate correction for multiple testing, with *p-value* thresholds of 0.01 and 0.005. Consistent TAD boundary calls with negative TAD separation scores were selected to define domain limits at 10 kb and 20 kb matrix resolutions.

CTCF motif binding predictions

The program *fimo* was used to scan for the *J. coenia* candidate TAD boundary region (HiC_scaffold_12:6430000-6444000) for canonical CTCF binding sites, using the positional weight matrix MA0205.1 deposited in the JASPAR database (Holohan et al. 2007 [↗](#); Cuellar-Partida et al. 2012 [↗](#); Castro-Mondragon et al. 2022 [↗](#)).

Genomic conservation analyses

We generated whole-genome alignments of 25 Lepidoptera and 2 Trichoptera reference species from NCBI Assembly using *ProgressiveCactus* (Armstrong et al. 2020 [↗](#)), and *HAL tools* (Hickey et al. 2013 [↗](#)) for converting the resulting HAL file to the MAF format. We provided a species

topology tree of 23 Lepidoptera species to *PhyloFit* (Hubisz et al. 2011 [DOI](#)) to fit a multiple sequence alignment on the reference *J. coenia* *Ubx* locus, using *HKY85* as the substitution model. Using *PhastCons* (Siepel et al. 2005 [DOI](#)), we then generated conservation score plots using standard parameters (target-coverage = 0.45; expected-length = 12; rho = 0.1) stored in BED and WIG file formats.

Butterfly rearing and CRISPR microinjections

J. coenia and *V. cardui* colonies were maintained at 25°C and 60-70% relative humidity in a growth chamber with a 14:10 light:dark photoperiod. Larval rearing on artificial diets, egg collection, and microinjections followed previously described methods (Martin et al. 2020 [DOI](#); Tendolkar et al. 2021 [DOI](#)). Cas9:sgRNA heteroduplexes were prepared as previously described (Martin et al., 2020 [DOI](#)). Frozen aliquots of Cas9-2xNLS (2.5 μ L ; 1,000 ng/ μ L) and sgRNA (2.5 μ L ; 500 ng/ μ L) were mixed in 2:1 and 4:1:1 mass ratios for single and dual target injections, respectively. CRISPR sgRNA targets are listed in **Table S1**.

Genotyping

For verification in DNA lesions at the intended Antp-Ubx_BE site, a pupal wing fragment harboring visible mutant clones (Fig. S4B”) or control wild-type tissue were PCR amplified using the diluted protocol of the Phire Animal Tissue Direct PCR Kit (ThermoFisher) and a pair of oligonucleotides (Forward : 5'-ACCGATCGTAAACGTCACCTTAACG-3' ; Reverse 5'-TACTGCGGTGGCGAGTGAATG-3'), before purification and Sanger sequencing using the reverse primer.

Antibody stainings

Fifth instar wing disks were dissected in ice cold Phosphate Buffer Saline (PBS), fixed for 15-20 min at room temperature in methanol-free formaldehyde diluted to 4% in PBS / 2 mM EGTA (egtazic acid), washed in PBS with 0.1% Triton X-100 (PT), stored in PT with 0.5% Bovine Serum albumin (PT-BSA), incubated overnight at 4°C in PT-BSA with a 1:5 dilution of the anti-UbdA peptide antibody serum (mouse monoclonal FP6.87, Developmental Studies Hybridoma Bank), and washed in PT. A 1:250 dilution of anti-Mouse IgG antibody coupled to AlexaFluor488 or Rabbit AlexaFluor555 was made in PT-BSA and spun down at 14,000 rcf to remove aggregates, and incubated with wings for 2 h at room temperature, before additional washes, incubation in 50% glycerol-PBS with DAPI (4',6-diamidino-2-phenylindole) nuclear stain, and incubation and mounting in 60% glycerol-PBS with 2 mM of EDTA (Ethylenediaminetetraacetic acid).

Imaging

Full-mount photographs of *J. coenia* and *V. cardui* were taken on a Nikon D5300 digital camera mounted with an AF-S VR MicroNikkor 105mm f/2.8G lens, with magnified views taken using a Keyence VHX-5000 digital microscope mounted with VH-Z00T and VH-Z100T lenses. Immunofluorescent stainings were imaged on an Olympus BX53 epifluorescent microscope mounted with UPLFLN 4x/0.13 and 10X/0.3 objectives.

Acknowledgements

We thank Ling Sheng Loh and the undergraduate researchers from the Martin Lab for assistance with micro-injections and insect rearing, Rachel Canalicchio and the GWU Harlan Greenhouse personnel for growing host plants, Patricia Hernandez for sharing microscopes, and Alex Wild for assistance with *Heliconius* microphotographs at UT Austin. We wish to acknowledge James Lewis and Bob Reed for stimulating insights on open-chromatin biology and the *Hox* locus, as well as for generating Hi-C libraries published in previous publications that we re-analyzed here. This work

was supported by the NSF awards IOS-1656553 and IOS-2110534 to AM, the Wilbur V. Harlan Research Fellowship to AT, the NSF Postdoctoral Research Fellowship in Biology to AMV, and the Smithsonian Institution Biodiversity Genomics Fellowship to JJH.

References

- Adikusuma F, Piltz S, Corbett MA, et al (2018) **Large deletions induced by Cas9 cleavage** *Nature* **560**:E8–E9
- Anania C, Acemel RD, Jedamzick J, et al (2022) **In vivo dissection of a clustered-CTCF domain boundary reveals developmental principles of regulatory insulation** *Nature Genetics* **54**:1026–1036
- Armstrong J, Hickey G, Diekhans M, et al (2020) **Progressive Cactus is a multiple-genome aligner for the thousand-genome era** *Nature* **587**:246–251 <https://doi.org/10.1038/s41586-020-2871-y>
- Arzate-Mejía RG, Josué Cerecedo-Castillo A, Guerrero G, et al (2020) **In situ dissection of domain boundaries affect genome topology and gene transcription in *Drosophila*** *Nature communications* **11**
- Bowman SK, Deaton AM, Domingues H, et al (2014) **H3K27 modifications define segmental regulatory domains in the *Drosophila* bithorax complex** *Elife* **3**
- Buffry AD, Kittelmann S, McGregor AP (2023) **Characterisation of the role and regulation of Ultrabithorax in sculpting fine-scale leg morphology** *Frontiers in Cell and Developmental Biology* **11**
- Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, et al (2022) **JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles** *Nucleic acids research* **50**:D165–D173
- Challis RJ, Kumar S, Dasmahapatra KKK, et al (2016) **Lepbase: the Lepidopteran genome database** *bioRxiv*
- Connahs H, Tlili S, van Creijl J, et al (2019) **Activation of butterfly eyespots by Distal-less is consistent with a reaction-diffusion process** *Development* **146**
- Cuellar-Partida G, Buske FA, McLeay RC, et al (2012) **Epigenetic priors for identifying active transcription factor binding sites** *Bioinformatics* **28**:56–62
- Cullot G, Boutin J, Toutain J, et al (2019) **CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations** *Nature communications* **10**
- Dobin A, Gingeras TR (2016) **Optimizing RNA-Seq mapping with STAR** *Data mining techniques for the life sciences* :245–262
- Feng S, Rastogi C, Loker R, et al (2022) **Transcription factor paralogs orchestrate alternative gene regulatory networks by context-dependent cooperation with multiple cofactors** *Nature Communications* **13**:1–19
- Ferguson L, Marletaz F, Carter J-M, et al (2014) **Ancient expansion of the Hox cluster in Lepidoptera generated four homeobox genes implicated in extra-embryonic tissue formation** *PLoS Genet* **10**

- Gambetta MC, Furlong EE (2018) **The insulator protein CTCF is required for correct Hox gene expression, but not for embryonic development in Drosophila** *Genetics* **210**:129–136
- Garaulet DL, Lai EC (2015) **Hox miRNA regulation within the Drosophila Bithorax complex: Patterning behavior** *Mechanisms of development* **138**:151–159
- Gaunt SJ (2022) **Seeking sense in the Hox gene cluster** *Journal of Developmental Biology* **10**
- Gisselbrecht SS, Palagi A, Kurland JV, et al (2020) **Transcriptional silencers in drosophila serve a dual role as transcriptional enhancers in alternate cellular contexts** *Molecular cell* **77**:324–337
- Gummalla M, Maeda RK, Castro Alvarez JJ, et al (2012) **abd-A regulation by the iab-8 noncoding RNA** *PLoS genetics* **8**
- Guo Y, Xu Q, Canzio D, et al (2015) **CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function** *Cell* **162**:900–910
- Gutierrez-Perez I, Rowley MJ, Lyu X, et al (2019) **Ecdysone-Induced 3D chromatin reorganization involves active enhancers bound by pipsqueak and polycomb** *Cell reports* **28**:2715–2727
- Hajirnis N, Mishra RK (2021) **Homeotic Genes: Clustering, Modularity, and Diversity** *Frontiers in Cell and Developmental Biology* **9**
- Hanly JJ, Livraghi L, Heryanto C, et al (2022) **A large deletion at the cortex locus eliminates butterfly wing patterning** *G3* **12**
- Hanly JJ, Wallbank RW, McMillan WO, Jiggins CD (2019) **Conservation and flexibility in the gene regulatory landscape of heliconiine butterfly wings** *Evodevo* **10**:1–14
- Heffer A, Pick L (2013) **Conservation and variation in Hox genes: how insect models pioneered the evo-devo field** *Annu Rev Entomol* **58**:161–179
- Hermann A, Kosman D, McGinnis W, Tour E (2022) **The expression of Drosophila melanogaster Hox gene Ultrabithorax is not overtly regulated by the intronic long noncoding RNA lncRNA:PS4 in a wild-type genetic background** *G3 Genes | Genomes | Genetics* **12** <https://doi.org/10.1093/g3journal/jkab374>
- Hickey G, Paten B, Earl D, et al (2013) **HAL: a hierarchical format for storing and analyzing multiple genome alignments** *Bioinformatics* **29**:1341–1342 <https://doi.org/10.1093/bioinformatics/btt128>
- Holohan EE, Kwong C, Adryan B, et al (2007) **CTCF genomic binding sites in Drosophila and the organisation of the bithorax complex** *PLoS genetics* **3**
- Hubisz MJ, Pollard KS, Siepel A (2011) **PHAST and RPHAST: phylogenetic analysis with space/time models** *Brief Bioinform* **12**:41–51 <https://doi.org/10.1093/bib/bbq072>
- Hughes CL, Kaufman TC (2002) **Hox genes and the evolution of the arthropod body plan** *Evolution & development* **4**:459–499
- Ibragimov A, Bing XY, Shidlovskii Y, et al (2023) **lncRNA read-through regulates the BX-C insulator Fub-1** *Elife* **12**

- Kahn TG, Savitsky M, Kuong C, et al (2023) **Topological screen identifies hundreds of Cp190- and CTCF-dependent Drosophila chromatin insulator elements** *Science Advances* **9**
- Kaushal A, Dorier J, Wang B, et al (2022) **Essential role of Cp190 in physical and regulatory boundary formation** *Science Advances* **8**
- Kawahara AY, Plotkin D, Espeland M, et al (2019) **Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths** *Proceedings of the National Academy of Sciences* **116**:22657–22663
- Kitts PA, Church DM, Thibaud-Nissen F, et al (2016) **Assembly: a resource for assembled genomes at NCBI** *Nucleic acids research* **44**:D73–D80
- Kosicki M, Tomberg K, Bradley A (2018) **Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements** *Nature biotechnology* **36**:765–771
- Kyrchanova O, Maksimenko O, Ibragimov A, et al (2020) **The insulator functions of the Drosophila polydactyl C2H2 zinc finger protein CTCF: Necessity versus sufficiency** *Science advances* **6**
- Lewis DL, DeCamillis MA, Brunetti CR, et al (1999) **Ectopic gene expression and homeotic transformations in arthropods using recombinant Sindbis viruses** *Current Biology* **9**:1279–1287 [https://doi.org/10.1016/S0960-9822\(00\)80049-4](https://doi.org/10.1016/S0960-9822(00)80049-4)
- Lewis EB (1978) **A gene complex controlling segmentation in Drosophila** *Nature* **276**:565–570
- Lewis JJ, Reed RD (2018) **Genome-wide regulatory adaptation shapes population-level genomic landscapes in Heliconius** *Molecular biology and evolution*
- Livraghi L (2017) **Hox3 duplication and divergence in the Lepidoptera** *PhD Thesis, Oxford Brookes University*
- Livraghi L, Martin A, Gibbs M, et al (2017) **CRISPR/Cas9 as the Key to Unlocking the Secrets of Butterfly Wing Pattern Development and Its Evolution** *Advances in Insect Physiology*
- Lohse K, Mackintosh A, Vila R, Consortium DT (2021) **The genome sequence of the European peacock butterfly *Aglais io* (Linnaeus, 1758)**. *Wellcome Open Research* **6**
- Lohse K, Wright C, Talavera G, et al (2021) **The genome sequence of the painted lady, *Vanessa cardui* Linnaeus 1758** *Wellcome Open Research* **6**
- Loker R, Sanner JE, Mann RS (2021) **Cell-type-specific Hox regulatory strategies orchestrate tissue identity** *Current Biology* **31**:4246–4255
- Love MI, Huber W, Anders S (2014) **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2** *Genome biology* **15**:1–21
- Lupiáñez DG, Kraft K, Heinrich V, et al (2015) **Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions** *Cell* **161**:1012–1025
- Maeda RK, Karch F (2015) **The open for business model of the bithorax complex in Drosophila** *Chromosoma* **124**:293–307

- Mahfooz N, Turchyn N, Mihajlovic M, et al (2007) **Ubx regulates differential enlargement and diversification of insect hind legs** *PLoS One* **2**
- Mallo M, Alonso CR (2013) **The regulation of Hox gene expression during animal development** *Development* **140**:3951–3963
- Martin A, Wolcott NS, O’Connell LA (2020) **Bringing immersive science to undergraduate laboratory courses using CRISPR gene knockouts in frogs and butterflies** *Journal of Experimental Biology* **223**
- Masumoto M, Yaginuma T, Niimi T (2009) **Functional analysis of Ultrabithorax in the silkworm, *Bombyx mori*, using RNAi** *Development genes and evolution* **219**:437–444
- Matsuoka Y, Monteiro A (2018) **Melanin pathway genes regulate color and morphology of butterfly wing scales** *Cell reports* **24**:56–65
- Matsuoka Y, Monteiro A (2021) **Hox genes are essential for the development of eyespots in *Bicyclus anynana* butterflies** *Genetics* **217**
- Matsuoka Y, Monteiro A (2022) **Ultrabithorax modifies a regulatory network of genes essential for butterfly eyespot development in a wing sector-specific manner** *Development* **149**
- Matsuoka Y, Murugesan SN, Prakash A, Monteiro A (2022) **Lepidopteran prolegs are novel traits, not leg homologs** *bioRxiv*
- Mazo-Vargas A, Langmüller AM, Wilder A, et al (2022) **Deep cis-regulatory homology of the butterfly wing pattern ground plan** *Science* **378**:304–308
- Méndez-González ID, Williams TM, Rebeiz M (2023) **Changes in locus wide repression underlie the evolution of *Drosophila* abdominal pigmentation** *PLoS genetics* **19**
- Merabet S, Carnesecchi J (2022) **Hox dosage and morphological diversification during development and evolution** *Seminars in Cell & Developmental Biology*
- Mitra S, Biswas A, Narlikar L (2018) **DIVERSITY in binding, regulation, and evolution revealed from high-throughput ChIP** *PLoS Computational Biology* **14**
- 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**
- Mulhair P, Crowley L, Boyes DH, et al (2022) **Diversity, duplication, and genomic organization of homeobox genes in Lepidoptera** *Genome Research* *gr*
- Mulhair PO, Holland PW (2022) **Evolution of the insect Hox gene cluster: Comparative analysis across 243 species** *Seminars in Cell & Developmental Biology*
- 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**
- Pease B, Borges AC, Bender W (2013) **Noncoding RNAs of the Ultrabithorax domain of the *Drosophila bithorax* complex** *Genetics* **195**:1253–1264
- Petruk S, Sedkov Y, Riley KM, et al (2006) **Transcription of bxd Noncoding RNAs Promoted by Trithorax Represses Ubx in cis by Transcriptional Interference** *Cell* **127**:1209–1221 <https://doi.org/10.1016/j.cell.2006.10.039>
- 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**
- 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**
- Ramírez F, Bhardwaj V, Arrigoni L, et al (2018) **High-resolution TADs reveal DNA sequences underlying genome organization in flies** *Nature communications* **9**
- 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**
- 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**
- 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**
- 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**
- 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**

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

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**

Zheng Z, Khoo A, Fambrough 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**:1650–16

Article and author information

Amruta Tendolkar

Department of Biological Sciences, The George Washington University, Washington, DC, United States

ORCID iD: [0000-0002-8496-3669](https://orcid.org/0000-0002-8496-3669)

Anyi Mazo-Vargas

Department of Biological Sciences, The George Washington University, Washington, DC, United States

ORCID iD: [0000-0001-9644-2871](https://orcid.org/0000-0001-9644-2871)

Luca Livraghi

Department of Biological Sciences, The George Washington University, Washington, DC, United States

ORCID iD: [0000-0002-2597-7550](https://orcid.org/0000-0002-2597-7550)

Joseph J. Hanly

Department of Biological Sciences, The George Washington University, Washington, DC, United States, Smithsonian Tropical Research Institute, Panama

ORCID iD: [0000-0002-9459-9776](https://orcid.org/0000-0002-9459-9776)

Kelsey C. Van Horne

Department of Biological Sciences, The George Washington University, Washington, DC, United States

Lawrence E. Gilbert

Department of Integrative Biology, University of Texas – Austin, Austin, TX, United States

ORCID iD: [0000-0003-4418-2996](https://orcid.org/0000-0003-4418-2996)

Arnaud Martin

Department of Biological Sciences, The George Washington University, Washington, DC, United States

For correspondence: arnaud@gwu.edu

ORCID iD: [0000-0002-5980-2249](https://orcid.org/0000-0002-5980-2249)

Copyright

© 2023, Tendolkar et al.

This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

Editors

Reviewing Editor

Ariel Chipman

The Hebrew University of Jerusalem, Jerusalem, Israel

Senior Editor

Claude Desplan

New York University, New York, United States of America

Joint Public Review:

Summary:

The existence of hox gene complexes conserved in animals with bilateral symmetry and in which the genes are arranged along the chromosome in the same order as the structures they specify along the anteroposterior axis of organisms is one of the most spectacular discoveries of recent developmental biology. In brief, homeotic mutations leads to the transformation of a given body segment of the fly into the copy of the next adjacent segment. For the sake of understanding the main observation of this work, it is important to know that in loss-of-function (LOF) alleles, a given segment develops like a copy of the segment immediately anterior to it, and in gain-of-function mutations (GOF), the affected segment develop like a copy of the immediately posterior segment. Over the last 30 years the molecular lesions associated with GOF alleles led to a model where the sequential activation of the hox genes along the chromosome result from the sequential opening of chromosomal domains. Most of these GOF alleles turned out to be deletions of boundary elements (BE) that define the extend

of the segment-specific regulatory domains. The fruit fly *Drosophila* is a highly specialized insect with a very rapid mode of segmentation. Furthermore, the *hox* clusters in this lineage have split. Given these specificities it is legitimate to question whether the regulatory landscape of the BX-C we know of in *D.melanogaster* is the result of very high specialization in this lineage, or whether it reflects a more ancestral organization. In this article, the authors address this question by analyzing the continuous *hox* cluster in butterflies. They focus on the intergenic region between the *Antennapedia* and the *Ubx* gene, where the split occurred in *D.melanogaster*. Hi-C and ATAC-seq data suggest the existence of a boundary element between 2 Topologically-Associated-Domain (TAD) which is also characterized by the presence of CTCF binding sites. Butterflies have 2 pairs of wings originating from T2 (forewing) specified by *Antp* and T3 specified by *Ubx* (hindwing). Remarkably, CRISPR mutational perturbation of this boundary leads to the hatching of butterflies with homeotic clones of cells with hindwings identities in the forewing (a posteriorly oriented homeotic transformation). In agreement with this phenotype, the authors observe ectopic expression of *Ubx* in these clones of cells. In other words, CRISPR mutagenesis of this BE region identified by molecular tool give rise to homeotic transformations directed towards more posterior segment as the boundary mutations that had been 1st identified on the basis of their posterior oriented homeotic transformation in *Drosophila*. None of the mutant clones they observed affect the hindwing, indicating that their scheme did not affect the nearby *Ubx* transcription unit. This is a reassuring and important 1st evidence that some of the regulatory paradigm that have been proposed in fruit flies are also at work in the common ancestor to *Drosophilae* and *Lepidoptera*.

Given the large size of the *Ubx* transcription unit and its associated regulatory regions it is not surprising that the authors have identified ncRNA that are conserved in 4 species of Nymphalinae butterflies, some of which also present in *D.melanogaster*. Attempts to target the promoters by CRISPR give rise to clones of cells in both forewings and hindwings, suggesting the generation of regulatory mutations associated with both LOF and GOF transformations. The presence of clones with dual homeosis suggest the targeting of *Ubx* activator and repression CRMs. Unfortunately, these experiments do not allow us to make further conclusions on the role of these ncRNA or in the identification of specific regulatory elements. To the opinion of this referee, some recent papers addressing the role that these ncRNA may play into boundary function should be taken with caution, and evidences that ncRNA(s) regulate boundaries in the BX-C in a WT context are still lacking.

Strengths: the convincing GOF phenotype resulting from the targeting of the *Antp-Ubx_BE*

Weaknesses: the lack of comparisons with the equivalent phenotypes obtained in *D.melanogaster* with for example the *Fub* mutation

- <https://doi.org/10.7554/eLife.90846.2.sa0>

Author Response

The following is the authors' response to the original reviews.

eLife assessment

This valuable paper examines the Bithorax complex in several butterfly species, in which the complex is contiguous and not split, as it is in the well-studied fruit fly Drosophila. Based on genetic screens and genetic manipulations of a boundary element involved in segment-specific regulation of Ubx, the authors provide solid evidence for their conclusions, which could be further strengthened by additional data and analyses. The data presented are relevant for those interested in the evolution and function of Hox genes and of gene regulation in general.

We are deeply grateful to the eLife editorial team and the two reviewers for their thoughtful and constructive feedback. We have used this feedback to improve our manuscript and have provided a point-by-point response below.

Public Reviews:

Reviewer #1 (Public Review):

In their article, "Cis-regulatory modes of Ultrabithorax inactivation in butterfly forewings," Tendolkar and colleagues explore Ubx regulation in butterflies. The authors investigated how Ubx expression is restricted to the hindwing in butterflies through a series of genomic analyses and genetic perturbations. The authors provide evidence that a Topologically Associated Domain (TAD) maintains a hindwing-enriched profile of chromatin around Ubx, largely through an apparent boundary element. CRISPR mutations of this boundary element led to ectopic Ubx expression in forewings, resulting in homeotic transformation in the wings. The authors also explore the results of the mutation in two non-coding RNA regions as well as a possible enhancer module. Each of these induces homeotic phenotypes. Finally, the authors describe a number of homeotic phenotypes in butterflies, which they relate to their work.

Together, this was an interesting paper with compelling initial data. That said, I have several items that I feel would warrant further discussion, presentation, or data.

First, I would not state, "Little is known about how Hox genes are regulated outside of flies." They should add "in insects" since so much is known in vertebrates

Corrected

For Figure 1, it would aid the readers if the authors could show the number of RNAseq reads across the locus. This would allow the readership to evaluate the frequency of the lncRNAs, splice variants, etc.

We have found it useful in the past to feature "Sashimi Plots", as they provide a good overview of transcript splicing junctions and read support. Here we could not accommodate this in our Fig. 1A as this would require compiling the RNAseq reads from many tissues and stages to be meaningful, and we would lose the resolution on forewing vs hindwing tissues that is important in this article (only the *Kallima inachus* dataset allows this comparison, and was used in Fig 1B). More specifically, the wing transcriptomes available for *J. coenia* and *V. cardui* are not deep enough to provide a good visualization of Antp alternative promoter usage or on AS5' transcription.

How common are boundary elements within introns? Typically, boundary elements are outside gene bodies, so this could be explored further. This seems like an interesting bit of biology which, following from the above point, it would be interesting to, at a minimum, discuss, but also relate to how transcription occurs through a possible boundary element (are there splice variants, for example?).

We do not see evidence of alternative splicing, and prefer to avoid speculating on transcriptional effects, but we agree that the intragenicity of the TAD boundary is interesting. We briefly highlighted this point in the revised Discussion:

"Lastly, it is worth noting that the Antp/Ubx TAD boundary we identified is intragenic, within the last intron of Ubx. It is unclear if this feature affects Ubx transcription, but this configuration might be analogue to the Notch locus in *Drosophila*, which includes a functional TAD boundary in an intronic position (Arzate-Mejía et al. 2020)."

The CRISPR experiments led to compelling phenotypes. However, as a Drosophila biologist, I found it hard to interpret the data from mosaic experiments. For example, in control experiments, how often do butterflies die? Are there offsite effects? It's striking that single-guide RNAs led to such strong effects. Is this common outside of this system? Is it possible to explore the function effects at the boundary element - are these generating large deletions (for example, like Mazo-Vargas et al., 2022)? For the mosaic experiments, how frequent are these effects in nature or captive stocks? Would it be possible to resequence these types of effects? At the moment, this data, while compelling, was hard to put into the context of the experiments above without understanding how common the effects are. Ideally, there would be resequencing of these tissues, which could be targeted, but it was not clear to me the general rates of these variants.

We agree with this assessment completely: mosaics complicate the proper interpretation of CRISPR based perturbation assays in regulatory regions. Here, unlike in Mazo-Vargas et al. (2022), we were unable to breed homeotic effects to a G1 generation, possibly because the phenotypes are dominant and lethal at the embryonic stage (see also our reply to Reviewer 2). This means that mosaic mutants are often survivors with clones of restricted size in the wing, and they are probably rare, but we are unable to meaningfully measure a mutation spectrum frequency (e.g. how often large deletions are generated). As mentioned in the first paragraph of our Discussion, we think that many of the phenotypes we observed (besides the Ubx GOF effects from the BE targeting) were confounded by alleles that could include large SVs. We aim to address these questions in an upcoming manuscript, at a locus where regulatory perturbation does not impact survival, including using germline mutants and unbiased genotyping (whole genome resequencing).

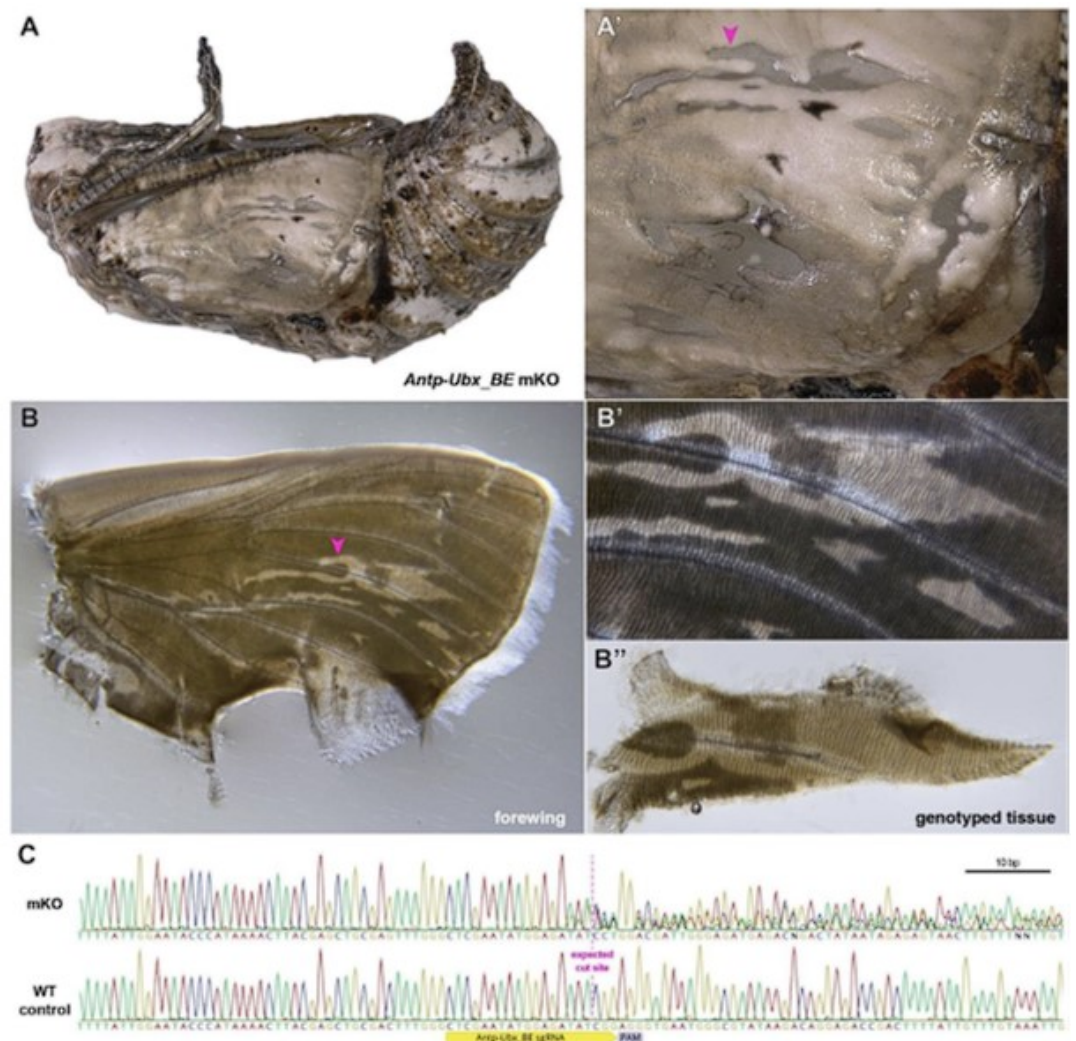
We elaborated on this issue in our Discussion:

"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). Ideally, regulatory CRISPR-induced alleles require genotyping in a second (G1) generation to be properly matched to a phenotype (Mazo-Vargas et al. 2022). Possibly because of lethal effects, we failed to pass G0 mutations to a G1 generation for genotyping, and were thus limited here to mosaic analysis. As adult wings have lost scale building cells that may underlie a given phenotype, we circumvented this issue by genotyping a pupal forewing displaying an homeotic phenotype in the more efficient Antp-Ubx_BE perturbation experiment (Fig. S4). In this case, PCR amplification of a 600 bp fragment followed by Sanger sequencing recovered signatures of indel variants, with mixed chromatograms starting at the targeted sites. But in all other experiments (CRM11, IT1, and AS5' targets), we did not genotype mutant tissues, as they were only detected in adult stages and generally with small clone sizes. Some of these clones may have been the results of large structural variants, as 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), and we are unable to assign a specific genotype to the homeotic effects we obtained in CRM11, IT1 and AS5' perturbation assays."

Our revision also includes a new Fig. S4 that features the mosaic genotyping of a G0 Antp-Ubx_BE mutant tissue. While this does not fully address the reviewer questions, it provides reasonable validation that the frequent GOF effects we observed upon perturbation at this target site are generated by on-target indels from DNA repair.

Author response image 1.

Validation of CRISPR-induced DNA Lesions in an Antp-Ubx_BE crispat pupat forewing. (A-A') Pupal forewing cuticle phenotype of an Antp-Ubx_BE J. coenia crispat, as in Fig. S3. (B-B'') Aspect of the same forewing under trans-illumination following dissection out of the pupal case. Regions from mutant clones have a more transparent appearance. (C). Sanger sequencing of an amplicon targeting the Antp-Ubx_BE region in the mutant tissue shown in panel B'', compared to a control wing tissue, showing mixed chromatogram around the expected CRISPR cutting site due to indel mutations from non-homologous end-joining.



In sum, I enjoyed the extensive mosaic perturbations. However, I feel that more molecular descriptions would elevate the work and make a larger impact on the field.

Reviewer #2 (Public Review):

Summary:

The existence of hox gene complexes conserved in animals with bilateral symmetry and in which the genes are arranged along the chromosome in the same order as the structures they specify along the anteroposterior axis of organisms is one of the most spectacular discoveries of recent developmental biology. In brief, homeotic mutations

lead to the transformation of a given body segment of the fly into a copy of the next adjacent segment. For the sake of understanding the main observation of this work, it is important to know that in loss-of-function (LOF) alleles, a given segment develops like a copy of the segment immediately anterior to it, and in gain-of-function mutations (GOF), the affected segment develops like a copy of the immediately posterior segment. Over the last 30 years the molecular lesions associated with GOF alleles led to a model where the sequential activation of the *hox* genes along the chromosome result from the sequential opening of chromosomal domains. Most of these GOF alleles turned out to be deletions of boundary elements (BE) that define the extent of the segment-specific regulatory domains. The fruit fly *Drosophila* is a highly specialized insect with a very rapid mode of segmentation. Furthermore, the *hox* clusters in this lineage have split. Given these specificities it is legitimate to question whether the regulatory landscape of the BX-C we know of in *D.melanogaster* is the result of very high specialization in this lineage, or whether it reflects a more ancestral organization. In this article, the authors address this question by analyzing the continuous *hox* cluster in butterflies. They focus on the intergenic region between the *Antennapedia* and the *Ubx* gene, where the split occurred in *D.melanogaster*. Hi-C and ATAC-seq data suggest the existence of a boundary element between 2 Topologically-Associated-Domain (TAD) which is also characterized by the presence of CTCF binding sites. Butterflies have 2 pairs of wings originating from T2 (forewing) specified by *Antp* and T3 specified by *Ubx* (hindwing). Remarkably, CRISPR mutational perturbation of this boundary leads to the hatching of butterflies with homeotic clones of cells with hindwings identities in the forewing (a posteriorly oriented homeotic transformation). In agreement with this phenotype, the authors observe ectopic expression of *Ubx* in these clones of cells. In other words, CRISPR mutagenesis of this BE region identified by molecular tool give rise to homeotic transformations directed towards more posterior segment as the boundary mutations that had been 1st identified on the basis of their posterior oriented homeotic transformation in *Drosophila*. None of the mutant clones they observed affect the hindwing, indicating that their scheme did not affect the nearby *Ubx* transcription unit. This is reassuring and important first evidence that some of the regulatory paradigms that have been proposed in fruit flies are also at work in the common ancestor to *Drosophilae* and *Lepidoptera*.

Given the large size of the *Ubx* transcription unit and its associated regulatory regions it is not surprising that the authors have identified ncRNA that are conserved in 4 species of *Nymphalinae* butterflies, some of which also present in *D.melanogaster*. Attempts to target the promoters by CRISPR give rise to clones of cells in both forewings and hindwings, suggesting the generation of regulatory mutations associated with both LOF and GOF transformations. The presence of clones with dual homeosis suggests the targeting of *Ubx* activator and repression CRMs. Unfortunately, these experiments do not allow us to make further conclusions on the role of these ncRNA or in the identification of specific regulatory elements. To the opinion of this reviewer, some recent papers addressing the role that these ncRNA may play in boundary function should be taken with caution, and evidence that ncRNA(s) regulate boundaries in the BX-C in a WT context is still lacking.

Strengths:

The convincing GOF phenotype resulting from the targeting of the *Antp-Ubx_BE*.

Weaknesses:

The lack of comparisons with the equivalent phenotypes obtained in *D.melanogaster* with for example the *Fub* mutation.

We are grateful for this excellent contextualization of our findings and have incorporated some of the historical elements into our revision, as detailed below.

Reviewer #2 (Recommendations For The Authors):

In the whole paper, the authors bring the notion of boundaries through the angle of the existence of TADs and ignore almost entirely to explain the characteristics of boundary mutation in the BX-C. To my knowledge examples where targeted boundary deletions between TADs result in misregulation of the neighboring genes, and/or a phenotype, are extremely sparse (especially in the context of the mouse hox genes). Given the extensive literature describing the boundary mutations and their associated GOF phenotypes, the paper would certainly gain strength if the authors justify their approach through this wealth of information. I must admit that this referee is surprised by the absence of any references to the founding work of the Karch and Bender laboratories on this topic. As a matter of fact, one of the founding members of the boundary class of regulatory elements was already brought in 1993 with the Fab-7 and Mcp elements of the BX-C. Based on gain-of-function homeotic phenotypes, additional Fab boundaries were added to the list. Finally, in 2013, Bender and Lucas (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3606092/>) identified the Fub boundary element that delimits the Ubx and abd-A domains in the BX-C. Fub fulfills the criterium of lying at the border of 2 neighboring TADs. Significantly, a deletion of Fub leads to a very penetrant and strong homeotic gain-of-function phenotype in which the flies hatch with a 1st abdominal segment transformed into the 2nd. In agreement with this, abd-A is expressed one parasegment too anterior in embryos. This is exactly the observation gathered from the targeted mutations in the Antp-Ubx_BE; a dominant transformation of anterior to posterior wing accompanied by an ectopic expression of Ubx in the forming primordia of the forwing where it is normally silenced. I believe the paper would gain credibility if the results were reported with the knowledge of the similarities with Fub.

Line 53, I am not aware of the existence of TADs for each of the 9 regulatory domains. The insulators delimit the extent of the regulatory domains but certainly not of TADs.

We thank the reviewer for these suggestions, as well as for the correction – we agree our previous text suggested that all BX-C boundaries are TAD boundaries, which was incorrect. We added a new introduction paragraph that combines classic literature on GOF mutations at boundary elements with recent evidence these are TAD insulators, including Fub (as suggested), and adding Fab-7 for breadth of scope.

"For instance, the deletion of a small region situated between Ubx and abd-A produces the Front-ultraabdominal phenotype (Fub) where the first abdominal segment (A1) is transformed into a copy of the second abdominal segment A2, due to a gain-of-expression of abd-A in A1 where it is normally repressed (Bender and Lucas 2013). At the molecular level, the Fub boundary is enforced by insulating factors that separate Topologically Associating Domains (TADs) of open-chromatin, while also allowing interactions of Ubx and abd-A enhancers with their target promoters (Postika et al. 2018; Srinivasan and Mishra 2020). Likewise, the Fab-7 deletion, which removes a TAD boundary insulating abd-A and Abd-B (Moniot-Perron et al. 2023), transforms parasegment 11 into parasegment 12 due to an anterior gain-of-expression of Abd-B (Gyurkovics et al. 1990). By extrapolation, one may expect that if the Drosophila Hox locus was not dislocated into two complexes, Antp and Ubx 3D contact domains would be separated by a Boundary Element (BE), and that deletions similar with Fub and Fab-7 mutations would result in gain-of-function mutations of Ubx that could effectively transform T2 regions into T3 identities."

A reference to the 1978 Nature article of Lewis should be added after line 42 of introduction.

Added

Line 56-57; the BX-C encoded miRNAs are known to regulate Ubx and abd-A, but not Abd-B.

Corrected

From lines 57 to 61, the authors mention reports aimed at demonstrating a role of ncRNA into Ubx regulation. To my eyes, these gathered evidences are rather weak. A reference to the work of Pease et al in Genetics in 2013 should be mentioned (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3832271/>).

Added. Our paragraph includes qualifier language about the functionality of the Ubx-related ncRNAs (“are thought to”, “appears to”), and updated references regarding bxd (Petruk et al. 2006; Ibragimov et al. 2023).

Line 62 authors, should write "Little is known about how Hox genes are regulated outside of Drosophila" and not flies.

Corrected

Lines 110-112 could lncRNA:Ubx-IT1 correspond to PS4 antisense reported by Pease et al in 2023 (see URL above)? Lines 115-117, could lncRNA:UbxAS5' correspond to bxd antisense of Pease et al in 2023 (see above)?

As we could not detect sequence similarities, we preferred to avoid drawing homology, and we intentionally avoided reference to the fly transcripts when we named IT1 and AS5'. This said, we agree it is important to clarify that further studies are needed to clarify this relationship. We elaborated on this point in our discussion:

"Of note, a systematic in-situ survey (Pease et al. 2013) showed that Drosophila embryos express an antisense transcripts in its 5' region (lncRNA:bxd), as well as within its first intron (lncRNA:PS4). It is thought that Drosophila bxd regulates Ubx, possibly by transcriptional interference or by facilitation of the Fub-1 boundary effect (Petruk et al. 2006; Ibragimov et al. 2023), while the possible regulatory roles of PS4 remain debated (Hermann et al. 2022). While these dipteran non-coding transcripts lack detectable sequence similarity with the lepidopteran IT1 and AS5' transcripts, further comparative genomics analyses of the Ubx region across the holometabolan insect phylogeny should clarify the extent to which Hox cluster lncRNAs have been conserved or independently evolved."

Lines 154-155: "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 ». This is only correlative, I would write "suggests" instead of "indicates" and add a "might function".

Corrected as suggested.

Line 254, I assume the authors wish to write Ubx-IT1 in V. cardui instead of Ubx-T1.

Typo corrected

Line 255 : Fig.5 is absent from the pdf file and replaced by table 1. I did not find a legend for Table 1.

Corrected, with our sincere apologies for the loss of this image in our first submission.

Line 293 "Individual with hindwing clones 2.75 times more common than...." "are" is missing?

Corrected

Lines 303-313, it is not entirely clear how many guide RNAs were injected. Would be useful to indicate the sites targeted in Fig.S8.

We specify in the revised text : using a single guide RNA (Ubx11b9)

Lines 323-337: it is not entirely clear to this referee (a drosophilist) if those spontaneous mutations can be inbred or whether these individuals are occasional mosaics. In general, did anyone try to derive lines from those mosaic animals? Is it possible to hit the germline at the syncytial stages at which the guides are injected? Are the individuals with wing phenotype fertile? Given the fact that the Antp-Ubx_BE mutations should be dominant, I wonder if this characteristic would not help in identifying germline transmission. Similar remark for the discussion where the authors explain at line 360, that genotyping can only be done in the progeny of the Go. I do not have the impression that the authors have performed this genotyping and if I am right, I do not understand why.

We improved our discussion section on this topic (new text in orange):

"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). Ideally, regulatory CRISPR-induced alleles require genotyping in a second (G1) generation to be properly matched to a phenotype (Mazo-Vargas et al. 2022). Possibly because of lethal effects, we failed to pass G0 mutations to a G1 generation for genotyping, and were thus limited here to mosaic analysis. As adult wings have lost scale building cells that may underlie a given phenotype, we circumvented this issue by genotyping a pupal forewing displaying an homeotic phenotype in the more efficient Antp-Ubx_BE perturbation experiment (Fig. S4). In this case, PCR amplification of a 600 bp fragment followed by Sanger sequencing recovered signatures of indel variants, with mixed chromatograms starting at the targeted sites. But in all other experiments (CRM11, IT1, and AS5' targets), we did not genotype mutant tissues, as they were only detected in adult stages and generally with small clone sizes. Some of these clones may have been the results of large structural variants, as 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), and we are unable to assign a specific genotype to the homeotic effects we obtained in CRM11, IT1 and AS5' perturbation assays."

We agree that the work we conducted with mosaics has important caveats. So far, our attempts at breeding homeotic G0 mutants have not been fruitful at this locus, while less deleterious loci can yield viable alleles into further generations, such as WntA (published) and cortex (in prep.). We prefer to stay vague about negative data here, as it is difficult to disentangle if they were due to real mutational effects (e.g. the alleles can be dominant and lethal in the G1 generation) to failure to germline carriers of mutations as founders, or to health issues that are often amplified by inbreeding depression (including a possible iflavirus in our *V. cardui* cultures).

We concur with the prediction that Antp-Ubx_BE mutations are probably dominant, and intend to follow up with similar GOF experiments in the *Plodia* pantry moth, a laboratory model for lepidopteran functional genomics that is more amenable than butterflies to inbreeding and long-term studies in mutant lines. In our experience (<https://www.frontiersin.org/articles/10.3389/fevo.2021.643661/full>), Ubx coding knock-out can be more extensive in *Plodia* than in butterflies, so we think these animals will also be more resilient to the deleterious effects of the GOF phenotype.

| *Line 423, 425, I am not a fan of the term "de-insulating!!!!*

We replaced this neologism by Similar deletion alleles resulting in a TAD fusion and misexpression effect (see below).

| *Line 425, why bring the work on Notch while there are so many examples in the BX-C itself....*

Our revised sentence makes it more clear we are referring here to documented examples of deletion-mediated TAD fusion (ie. featuring a conformation capture assay such as HiC/micro-C):

| *This suggests a possible loss 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 deletion alleles resulting in a TAD fusion and misexpression effect 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).*

Our revision also includes more emphasis on the *Drosophila* BX-C boundary elements Fub and Fab-7 (see above).